

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 April 2003 (17.04.2003)

PCT

(10) International Publication Number
WO 03/031588 A2

(51) International Patent Classification⁷: C12N

(21) International Application Number: PCT/US02/32512

(22) International Filing Date: 10 October 2002 (10.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/328,655 11 October 2001 (11.10.2001) US
60/363,774 13 March 2002 (13.03.2002) US

(71) Applicants (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI, S.P.A. [IT/IT]; VIA PONTINA KM. 30.600, I-00040 POMEZIA (IT).

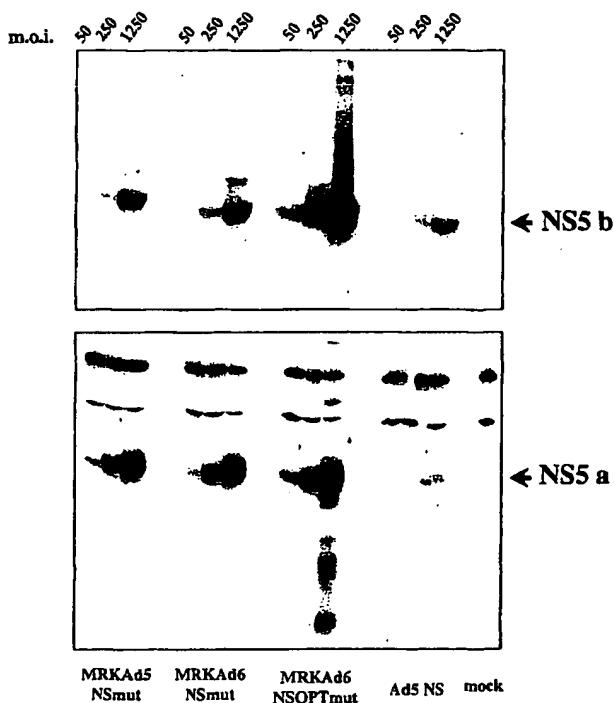
(72) Inventors; and

(75) Inventors/Applicants (for US only): EMINI, Emilio, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). KASLOW, David, C. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). BETT, Andrew, J. [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). NICOSIA, Alfredo [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LAHM, Armin [DE/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LUZZAGO, Alessandra [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). CORTESE, Riccardo [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). COLLOCA, Stefano [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT).

(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

[Continued on next page]

(54) Title: HEPATITIS C VIRUS VACCINE



(57) Abstract: The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

WO 03/031588 A2

Western blot on whole-cell extracts from HeLa cells infected at different multiplicity of infection (m.o.i.; indicated at the top) with Adenovectors expressing the different HCV NS cassettes. Mature NS5B and NS5A products were detected with specific antibodies.



(81) **Designated States (national):** AE, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CII, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI, GM, IIR, IIU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PII, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

European patent (AT, BE, BG, CII, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION
HEPATITIS C VIRUS VACCINE

RELATED APPLICATIONS

5 The present application claims priority to provisional applications U.S. Serial No. 60/363,774, filed March 13, 2002, and U.S. Serial No. 60/328,655, filed October 11, 2001, each of which are hereby incorporated by reference herein.

BACKGROUND OF THE INVENTION

10 The references cited in the present application are not admitted to be prior art to the claimed invention.

About 3% of the world's population are infected with the Hepatitis C virus (HCV). (Wasley *et al.*, *Semin. Liver Dis.* 20, 1-16, 2000.) Exposure to HCV results in an overt acute disease in a small percentage of cases, while in most instances the virus establishes a chronic infection causing liver inflammation and slowly progresses into liver failure and cirrhosis. (Iwarson, *FEMS Microbiol. Rev.* 14, 201-204, 1994.) In addition, epidemiological surveys indicate an important role of HCV in the pathogenesis of hepatocellular carcinoma. (Kew, *FEMS Microbiol. Rev.* 14, 211-220, 1994, Alter, *Blood* 85, 1681-1695, 1995.)

20 Prior to the implementation of routine blood screening for HCV in 1992, most infections were contracted by inadvertent exposure to contaminated blood, blood products or transplanted organs. In those areas where blood screening of HCV is carried out, HCV is primarily contracted through direct percutaneous exposure to infected blood, *i.e.*, intravenous drug use. Less frequent methods of transmission include perinatal exposure, hemodialysis, and sexual contact with an HCV infected person. (Alter *et al.*, *N. Engl. J. Med.* 341(8), 556-562, 1999, Alter, *J. Hepatol.* 31 *Suppl.* 88-91, 1999. *Semin. Liver. Dis.* 201, 1-16, 2000.)

30 The HCV genome consists of a single strand RNA about 9.5 kb encoding a precursor polyprotein of about 3000 amino acids. (Choo *et al.*, *Science* 244, 362-364, 1989, Choo *et al.*, *Science* 244, 359-362, 1989, Takamizawa *et al.*, *J. Virol.* 65, 1105-1113, 1991.) The HCV polyprotein contains the viral proteins in the order: C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B.

Individual viral proteins are produced by proteolysis of the HCV polyprotein. Host cell proteases release the putative structural proteins C, E1, E2, and

p7, and create the N-terminus of NS2 at amino acid 810. (Mizushima *et al.*, *J. Virol.* 68, 2731-2734, 1994, Hijikata *et al.*, *P.N.A.S. USA* 90, 10773-10777, 1993.)

- The non-structural proteins NS3, NS4A, NS4B, NS5A and NS5B presumably form the virus replication machinery and are released from the
- 5 polyprotein. A zinc-dependent protease associated with NS2 and the N-terminus of NS3 is responsible for cleavage between NS2 and NS3. (Grakoui *et al.*, *J. Virol.* 67, 1385-1395, 1993, Hijikata *et al.*, *P.N.A.S. USA* 90, 10773-10777, 1993.) A distinct serine protease located in the N-terminal domain of NS3 is responsible for proteolytic cleavages at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/NS5B junctions.
- 10 (Bartenschlager *et al.*, *J. Virol.* 67, 3835-3844, 1993, Grakoui *et al.*, *Proc. Natl. Acad. Sci. USA* 90, 10583-10587, 1993, Tomei *et al.*, *J. Virol.* 67, 4017-4026, 1993.) NS4A provides a cofactor for NS3 activity. (Failla *et al.*, *J. Virol.* 68, 3753-3760, 1994, De Francesco *et al.*, U.S. Patent No. 5,739,002.)

15 NS5A is a highly phosphorylated protein conferring interferon resistance. (De Francesco *et al.*, *Semin. Liver Dis.*, 20(1), 69-83, 2000, Pawlotsky, *Viral Hepat. Suppl. 1*, 47-48, 1999.)

NS5B provides an RNA-dependent RNA polymerase. (De Francesco *et al.*, International Publication Number WO 96/37619, Behrens *et al.*, *EMBO* 15, 12-22, 1996, Lohmann *et al.*, *Virology* 249, 108-118, 1998.)

20

SUMMARY OF THE INVENTION

The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

A HCV specific CMI response refers to the production of cytotoxic T lymphocytes and T helper cells that recognize an HCV antigen. The CMI response may also include non-HCV specific immune effects.

Preferred nucleic acids encode a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide that is substantially similar to SEQ. ID. NO. 1 and has sufficient protease activity to process itself to produce at least a polypeptide substantially similar to the NS5B region present in SEQ. ID. NO. 1. The produced polypeptide corresponding to NS5B is enzymatically inactive. More preferably, the HCV polypeptide has sufficient

protease activity to produce polypeptides substantially similar to the NS3, NS4A, NS4B, NS5A, and NS5B regions present in SEQ. ID. NO. 1.

Reference to a "substantially similar sequence" indicates an identity of at least about 65% to a reference sequence. Thus, for example, polypeptides having an amino acid sequence substantially similar to SEQ. ID. NO. 1 have an overall amino acid identity of at least about 65% to SEQ. ID. NO. 1.

Polypeptides corresponding to NS3, NS4A, NS4B, NS5A, and NS5B have an amino acid sequence identity of at least about 65% to the corresponding region in SEQ. ID. NO. 1. Such corresponding polypeptides are also referred to 10 herein as NS3, NS4A, NS4B, NS5A, and NS5B polypeptides.

Thus, a first aspect of the present invention describes a nucleic acid comprising a nucleotide sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The encoded polypeptide has sufficient protease activity to process itself to produce an NS5B polypeptide that is 15 enzymatically inactive.

In a preferred embodiment, the nucleic acid is an expression vector capable of expressing the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide in a desired human cell. Expression inside a human cell has therapeutic applications for actively treating an HCV infection and for prophylactically treating against an HCV 20 infection.

An expression vector contains a nucleotide sequence encoding a polypeptide along with regulatory elements for proper transcription and processing. The regulatory elements that may be present include those naturally associated with the nucleotide sequence encoding the polypeptide and exogenous regulatory elements 25 not naturally associated with the nucleotide sequence. Exogenous regulatory elements such as an exogenous promoter can be useful for expression in a particular host, such as in a human cell. Examples of regulatory elements useful for functional expression include a promoter, a terminator, a ribosome binding site, and a polyadenylation signal.

Another aspect of the present invention describes a nucleic acid 30 comprising a gene expression cassette able to express in a human cell a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The polypeptide can process itself to produce an enzymatically inactive NS5B protein. The gene expression cassette contains at least the following:

- a) a promoter transcriptionally coupled to a nucleotide sequence encoding a polypeptide;
 - b) a 5' ribosome binding site functionally coupled to the nucleotide sequence,
- 5 c) a terminator joined to the 3' end of the nucleotide sequence, and
- d) a 3' polyadenylation signal functionally coupled to the nucleotide sequence.

Reference to "transcriptionally coupled" indicates that the promoter is positioned such that transcription of the nucleotide sequence can be brought about by RNA polymerase binding at the promoter. Transcriptionally coupled does not require that the sequence being transcribed is adjacent to the promoter.

Reference to "functionally coupled" indicates the ability to mediate an effect on the nucleotide sequence. Functionally coupled does not require that the coupled sequences be adjacent to each other. A 3' polyadenylation signal functionally coupled to the nucleotide sequence facilitates cleavage and polyadenylation of the transcribed RNA. A 5' ribosome binding site functionally coupled to the nucleotide sequence facilitates ribosome binding.

In preferred embodiments the nucleic acid is a DNA plasmid vector or an adenovector suitable for either therapeutic application in treating HCV or as an intermediate in the production of a therapeutic vector. Treating HCV includes actively treating an HCV infection and prophylactically treating against an HCV infection.

Another aspect of the present invention describes an adenovector comprising a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette able to express a polypeptide substantially similar to SEQ. ID. NO. 1 that is produced by a process involving (a) homologous recombination and (b) adenovector rescue. The homologous recombinant step produces an adenovirus genome plasmid. The adenovector rescue step produces the adenovector from the adenogenome plasmid.

Adenovirus genome plasmids described herein contain a recombinant adenovirus genome having a deletion in the E1 region and optionally in the E3 region and a gene expression cassette inserted into one of the deleted regions. The recombinant adenovirus genome is made of regions substantially similar to one or more adenovirus serotypes.

Another aspect of the present invention describes an adenovector consisting of the nucleic acid sequence of SEQ. ID. NO. 4 or a derivative thereof,

wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ. ID. NO. 4 replaced with the HCV polyprotein encoding sequence of either SEQ. ID. NO. 3, SEQ. ID. NO. 10 or SEQ. ID. NO. 11.

- Another aspect of the present invention describes a cultured
5 recombinant cell comprising a nucleic acid containing a sequence encoding a Met-
NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1.
The recombinant cell has a variety of uses such as being used to replicate nucleic acid
encoding the polypeptide in vector construction methods.

- Another aspect of the present invention describes a method of making
10 10 an adenovector comprising a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette
able to express a polypeptide substantially similar to SEQ. ID. NO. 1. The method
involves the steps of (a) producing an adenovirus genome plasmid containing a
recombinant adenovirus genome with deletions in the E1 and E3 regions and a gene
expression cassette inserted into one of the deleted regions and (b) rescuing the
15 adenovector from the adenovirus genome plasmid.

- Another aspect of the present invention describes a pharmaceutical
composition comprising a vector for expressing a Met-NS3-NS4A-NS4B-NS5A-
NS5B polypeptide substantially similar to SEQ. ID. NO. 1 and a pharmaceutically
acceptable carrier. The vector is suitable for administration and polypeptide
20 expression in a patient.

- A "patient" refers to a mammal capable of being infected with HCV.
A patient may or may not be infected with HCV. Examples of patients are humans
and chimpanzees.

- Another aspect of the present invention describes a method of treating
25 25 a patient comprising the step of administering to the patient an effective amount of a
vector expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially
similar to SEQ. ID. NO. 1. The vector is suitable for administration and polypeptide
expression in the patient.

- The patient undergoing treatment may or may not be infected with
30 30 HCV. For a patient infected with HCV, an effective amount is sufficient to achieve
one or more of the following effects: reduce the ability of HCV to replicate, reduce
HCV load, increase viral clearance, and increase one or more HCV specific CMI
responses. For a patient not infected with HCV, an effective amount is sufficient to
achieve one or more of the following: an increased ability to produce one or more
35 components of a HCV specific CMI response to a HCV infection, a reduced

susceptibility to HCV infection, and a reduced ability of the infecting virus to establish persistent infection for chronic disease.

Another aspect of the present invention features a recombinant nucleic acid comprising an Ad6 region and a region not present in Ad6. Reference to 5 "recombinant" nucleic acid indicates the presence of two or more nucleic acid regions not naturally associated with each other. Preferably, the Ad6 recombinant nucleic acid contains Ad6 regions and a gene expression cassette coding for a polypeptide heterologous to Ad6.

Other features and advantages of the present invention are apparent 10 from the additional descriptions provided herein including the different examples. The provided examples illustrate different components and methodology useful in practicing the present invention. The examples do not limit the claimed invention. Based on the present disclosure the skilled artisan can identify and employ other components and methodology useful for practicing the present invention.

15

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B illustrate SEQ. ID. NO. 1.

Figures 2A, 2B, 2C, and 2D illustrate SEQ. ID. NO. 2. SEQ. ID. NO. 20 provides a nucleotide sequence coding for SEQ. ID. NO. 1 along with an optimized internal ribosome entry site and TAAA termination. Nucleotides 1-6 provides an optimized internal ribosome entry site. Nucleotides 7-5961 code for a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide with nucleotides in positions 5137 to 5145 providing a AlaAlaGly sequence in amino acid positions 1711 to 1713 that renders NS5B inactive. Nucleotides 5962-5965 provide a TAAA termination.

Figures 3A, 3B, 3C, and 3D illustrate SEQ. ID. NO. 3. SEQ. ID. NO. 25 3 is a codon optimized version of SEQ. ID. NO. 2. Nucleotides 7-5961 encode a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

Figures 4A-4M illustrate MRKAd6-NSmut (SEQ. ID. NO. 4). SEQ. ID. NO. 4 is an adenovector containing an expression cassette where the polypeptide 30 of SEQ. ID. NO. 1 is encoded by SEQ. ID. NO. 2. Base pairs 1-450 correspond to the Ad5 bp 1 to 450; base pairs 462 to 1252 correspond to the human CMV promoter; base pairs 1258 to 1267 correspond to the Kozak sequence; base pairs 1264 to 7222 correspond to the NS genes; base pairs 7231 to 7451 correspond to the BGH polyadenylation signal; base pairs 7469 to 9506 correspond to Ad5 base pairs 3511 to 35 3548; base pairs 9507 to 32121 correspond to Ad6 base pairs 5542 to 28156; base

pairs 32122 to 35117 correspond to Ad6 base pairs 30789 to 33784; and base pairs 35118 to 37089 correspond to Ad5 base pairs 33967 to 35935.

- Figures 5A-5O illustrate SEQ. ID. NOs. 5 and 6. SEQ. ID. NO. 5 encodes a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide with an active 5 RNA dependent RNA polymerase. SEQ. ID. NO. 6 provides the amino acid sequence for the polypeptide.

Figures 6A-6C provide the nucleic acid sequence for pV1JnsA (SEQ. ID. NO. 7).

- Figures 7A-7N provide the nucleic acid sequence for the Ad6 genome 10 (SEQ. ID. NO. 8).

Figures 8A-8K provide the nucleic acid sequence for the Ad5 genome (SEQ. ID. NO. 9).

- Figure 9 illustrates different regions of the Ad6 genome. The linear (35759 bp) ds DNA genome is indicated by two parallel lines and is divided into 100 15 map units. Transcription units are shown relative to their position and orientation in the genome. Early genes (E1A, E1B, E2A/B, E3 and E4 are indicated by gray arrows. Late genes (L1 to L5), indicated by black arrows, are produced by alternative splicing of a transcript produced from the major late promoter (MLP) and all contain the tripartite leader (1, 2, 3) at their 5' ends. The E1 region is located from approximately 20 1.0 to 11.5 map units, the E2 region from 75.0 to 11.5 map units, E3 from 76.1 to 86.7 map units, and E4 from 99.5 to 91.2 map units. The major late transcription unit is located between 16.0 and 91.2 map units.

Figure 10 illustrates homologous recombination to recover pAdE1-E3+ containing Ad6 and Ad5 regions.

- 25 Figure 11 illustrates homologous recombinant to recover a pAdE1-E3+ containing Ad6 regions.

- Figure 12 illustrates a western blot on whole-cell extracts from 293 cells transfected with plasmid DNA expressing different HCV NS cassettes. Mature 30 NS3 and NSSA products were detected with specific antibodies. "pV1Jns-NS" refers to a pV1JnsA plasmid where a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is encoded by SEQ. ID. NO. 5, and SEQ. ID. NO. 5 is inserted between bases 1881 and 1912 of SEQ. ID. NO. 7. "pV1Jns-NSmut" refers to a pV1JnsA plasmid where SEQ. ID. NO. 2 is inserted between bases 1882 and 1925 of SEQ. ID. NO. 7. "pV1Jns-NSOPTmut" refers to a pV1JnsA plasmid where SEQ. ID. NO. 3 is inserted between 35 bases 1881 and 1905 of SEQ. ID. NO. 7.

Figures 13A and 13B illustrate T cell responses by IFN γ ELIspot induced in C57black6 mice (A) and BalbC mice (B) by two injections of 25 μ g and 50 μ g, respectively, of plasmid DNA encoding the different HCV NS cassettes with Gene Electro-Transfer (GET).

- 5 Figure 14 illustrates protein expression from different adenovectors upon infection of HeLa cells. MRKAd5-NSmut is an adenovector based on an Ad5 sequence (SEQ. ID. NO. 9), where the Ad5 genome has an E1 deletion of base pairs 451 to 3510, an E3 deletion of base pairs 28134 to 30817, and has the NS3-NS4A-NS4B-NS5A-NS5B expression cassette as provided in base pairs 451 to 7468 of SEQ.
10 ID. NO. 4 inserted between positions 450 and 3511. Ad5-NS is an adenovector based on an Ad5 backbone with an E1 deletion of base pairs 342 to 3523, and E3 deletion of base pairs 28134 to 30817 and containing an expression cassette encoding a NS3-NS4A-NS4B-NS5A-NS5B from SEQ. ID. NO. 5. "MRKAd6-NSOPTmut" refers to an adenovector having a modified SEQ. ID. NO. 4 sequence, wherein base pairs 1258 to 7222 of SEQ. ID. NO. 4 is replaced with SEQ. ID. NO. 3.

15 Figure 15 illustrates T cell responses by IFN γ ELIspot induced in C57black6 mice by two injections of 10⁹ vp of adenovectors containing different HCV non-structural gene cassettes.

- 20 Figures 16A-16D illustrate T cell responses by IFN γ ELIspot induced in Rhesus monkeys by one or two injections of 10¹⁰ vp (A) or 10¹¹ vp (B) of adenovectors containing different HCV non-structural gene cassettes.

25 Figures 17A and 17B illustrates CD8+ T cell responses by IFN γ ICS induced in Rhesus monkeys by two injections of 10¹⁰ vp (A) or 10¹¹ vp (B) of adenovectors encoding the different HCV non-structural gene cassettes.

- 25 Figures 18A-18F illustrate T cell responses by bulk CTL assay induced in Rhesus monkeys by two injections of 10¹¹ vp of Ad5-NS (A), MRKAd5-NSmut (B), or MRKAd6-NSmut (C).

 Figure 19 illustrates the plasmid pE2.

- 30 Figures 20A-D illustrates the partial codon optimized sequence NSsuboptmut (SEQ. ID. NO. 10). Coding sequence for the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is from base 7 to 5961.

DETAILED DESCRIPTION OF THE INVENTION

- The present invention features Ad6 vectors and nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide that contains an inactive NS5B region. Providing an inactive NS5B region supplies NS5B antigens while reducing the 5 possibility of adverse side effects due to an active viral RNA polymerase. Uses of the featured nucleic acid include use as a vaccine component to introduce into a cell an HCV polypeptide that provides a broad range of antigens for generating a CMI response against HCV, and as an intermediate for producing such a vaccine component.
- 10 The adaptive cellular immune response can function to recognize viral antigens in HCV infected cells throughout the body due to the ubiquitous distribution of major histocompatibility complex (MHC) class I and II expression, to induce immunological memory, and to maintain immunological memory. These functions are attributed to antigen-specific CD4+ T helper (Th) and CD8+ cytotoxic T cells (CTL).
- 15 Upon activation via their specific T cell receptors, HCV specific Th cells fulfill a variety of immunoregulatory functions, most of them mediated by Th1 and Th2 cytokines. HCV specific Th cells assist in the activation and differentiation of B cells and induction and stimulation of virus-specific cytotoxic T cells. Together 20 with CTL, Th cells may also secrete IFN- γ and TNF- α that inhibit replication and gene expression of several viruses. Additionally, Th cells and CTL, the main effector cells, can induce apoptosis and lysis of virus infected cells.
- 25 HCV specific CTL are generated from antigens processed by professional antigen presenting cells (pAPCs). Antigens can be either synthesized within or introduced into pAPCs. Antigen synthesis in a pAPC can be brought about by introducing into the cell an expression cassette encoding the antigen.
- 30 A preferred route of nucleic acid vaccine administration is an intramuscular route. Intramuscular administration appears to result in the introduction and expression of nucleic acid into somatic cells and pAPCs. HCV antigens produced in the somatic cells can be transferred to pAPCs for presentation in the context of MHC class I molecules. (Donnelly *et al.*, *Annu. Rev. Immunol.* 15:617-648, 1997.)
- pAPCs process longer length antigens into smaller peptide antigens in the proteasome complex. The antigen is translocated into the endoplasmic reticulum/Golgi complex secretory pathway for association with MHC class I

proteins. CD8+ T lymphocytes recognize antigen associated with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein.

- Using a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide as a vaccine component allows for production of a broad range of antigens capable of generating CMI responses from a single vector. The polypeptide should be able to process itself sufficiently to produce at least a region corresponding to NS5B. Preferred nucleic acids encode an amino acid sequence substantially similar to SEQ. ID. NO. 1 that has sufficient protease activity to process itself to produce individual HCV polypeptides substantially similar to the NS3, NS4A, NS4B, NS5A, and NS5B regions present in SEQ. ID. NO. 1.

A polypeptide substantially similar to SEQ. ID. NO. 1 with sufficient protease activity to process itself in a cell provides the cell with T cell epitopes that are present in several different HCV strains. Protease activity is provided by NS3 and NS3/NS4A proteins digesting the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide at the appropriate cleavage sites to release polypeptides corresponding to NS3, NS4A, NS4B, NS5A, and NS5B. Self-processing of the Met-NS3-NS4A-NS4B-NS5A-NS5B generates polypeptides that approximate naturally occurring HCV polypeptides.

Based on the guidance provided herein a sufficiently strong immune response can be generated to achieve beneficial effects in a patient. The provided guidance includes information concerning HCV sequence selection, vector selection, vector production, combination treatment, and administration.

I. HCV SEQUENCES

A variety of different nucleic acid sequences can be used as a vaccine component to supply a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide to a cell or as an intermediate to produce vaccine components. The starting point for obtaining suitable nucleic acid sequences are preferably naturally occurring NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequences modified to produce an inactive NS5B.

The use of a HCV nucleic acid sequence providing HCV non-structural antigens to generate a CMI response is mentioned by Cho *et al.*, *Vaccine* 17:1136-1144, 1999, Paliard *et al.*, International Publication Number WO 01/30812 (not admitted to be prior art to the claimed invention), and Coit *et al.*, International Publication Number WO 01/38360 (not admitted to be prior art to the claimed invention). Such references fail to describe, for example, a polypeptide that processes

itself to produce an inactive NS5B, and the particular combinations of HCV sequences and delivery vehicles employed herein.

Modifications to a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequence can be produced by altering the encoding nucleic acid.

5 Alterations can be performed to create deletions, insertions and substitutions.

Small modifications can be made in NS5B to produce an inactive polymerase by targeting motifs essentially for replication. Examples of motifs critical for NS5B activity and modifications that can be made to produce an inactive NS5B are described by Lohmann *et al.*, *Journal of Virology* 71:8416-8426, 1997, and

10 Kolykhalov *et al.*, *Journal of Virology* 74:2046-2051, 2000.

Additional factors to take into account when producing modifications to a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide include maintaining the ability to self-process and maintaining T cell antigens. The ability of the HCV polypeptide to process itself is determined to a large extent by a functional NS3

15 protease. Modifications that maintain NS3 activity protease activity can be obtained by taking into account the NS3 protein, NS4A which serves as a cofactor for NS3, and NS3 protease recognition sites present within the NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

Different modifications can be made to naturally occurring NS3-
20 NS4A-NS4B-NS5A-NS5B polypeptide sequences to produce polypeptides able to elicit a broad range of T cell responses. Factors influencing the ability of a polypeptide to elicit a broad T cell response include the preservation or introduction of HCV specific T cell antigen regions and prevalence of different T cell antigen regions in different HCV isolates.

25 Numerous examples of naturally occurring HCV isolates are well known in the art. HCV isolates can be classified into the following six major genotypes comprising one or more subtypes: HCV-1/(1a,1b,1c), HCV-2/(2a,2b,2c), HCV-3/(3a,3b,10a), HCV-4/(4a), HCV-5/(5a) and HCV-6/(6a,6b,7b,8b,9a,11a). (Simmonds, *J. Gen. Virol.*, 693-712, 2001.) Examples of particular HCV sequences
30 such as HCV-BK, HCV-J, HCV-N, HCV-H, have been deposited in GenBank and described in various publications. (See, for example, Chamberlain *et al.*, *J. Gen. Virol.*, 1341-1347, 1997.)

35 HCV T cell antigens can be identified by, for example, empirical experimentation. One way of identifying T cell antigens involves generating a series of overlapping short peptides from a longer length polypeptide and then screening the

T-cell populations from infected patients for positive clones. Positive clones are activated/primed by a particular peptide. Techniques such as IFN γ -ELISPOT, IFN γ -Intracellular staining and bulk CTL assays can be used to measure peptide activity. Peptides thus identified can be considered to represent T-cell epitopes of the
5 respective pathogen.

HCV T cell antigen regions from different HCV isolates can be introduced into a single sequence by, for example, producing a hybrid NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing regions from two or more naturally occurring sequences. Such a hybrid can contain additional modifications, which
10 preferably do not reduce the ability of the polypeptide to produce an HCV CMI response.

The ability of a modified Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide to process itself and produce a CMI response can be determined using techniques described herein or well known in the art. Such techniques include the use
15 of IFN γ -ELISPOT, IFN γ -Intracellular staining and bulk CTL assays to measure a HCV specific CMI response.

A. Met-NS3-NS4A-NS4B-NS5A-NS5B Sequences

SEQ. ID. NO. 1 provides a preferred Met-NS3-NS4A-NS4B-NS5A-NS5B sequence. SEQ. ID. NO. 1 contains a large number of HCV specific T cell antigens that are present in several different HCV isolates. SEQ. ID. NO. 1 is similar to the NS3-NS4A-NS4B-NS5A-NS5B portion of the HCV BK strain nucleotide sequence (GenBank accession number M58335).

In SEQ. ID. NO. 1 anchor positions important for recognition by MHC class I molecules are conserved or represent conservative substitutions for 18 out of 20 known T-cell epitopes in the NS3-NS4A-NS4B-NS5A-NS5B portion of HCV polyproteins. With respect to the remaining two known T-cell epitopes, one has a non-conservative anchor substitution in SEQ. ID. NO. 1 that may still be recognized by a different HLA supertype and one epitope has one anchor residue not conserved.
25 HCV T-cell epitopes are described in Chisari *et al.*, *Curr. Top. Microbiol. Immunol.*, 242:299-325, 2000, and Lechner *et al.* *J. Exp. Med.* 9:1499-1512, 2000.

Differences between the HCV-BK NS3-NS4A-NS4B-NS5A-NS5B nucleotide sequence and SEQ. ID. NO. 1 include the introduction of a methionine at the 5' end and the presence of modified NS5B active site residues in SEQ. ID. NO. 1.

The modification replaces GlyAspAsp with AlaAlaGly (residues 1711-1713) to inactivate NS5B.

The encoded HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide preferably has an amino acid sequence substantially similar to SEQ. ID. NO. 1. In

5 different embodiments, the encoded HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide has an amino acid identify to SEQ. ID. NO. 1 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or differs from SEQ. ID. NO. 1 by 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, or 1-20 amino acids.

10 Amino acid differences between a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide and SEQ. ID. NO. 1 are calculated by determining the minimum number of amino acid modifications in which the two sequences differ. Amino acid modifications can be deletions, additions, substitutions or any combination thereof.

15 Amino acid sequence identity is determined by methods well known in the art that compare the amino acid sequence of one polypeptide to the amino acid sequence of a second polypeptide and generate a sequence alignment. Amino acid identity is calculated from the alignment by counting the number of aligned residue pairs that have identical amino acids.

Methods for determining sequence identity include those described by 20 Schuler, G.D. in *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Baxevanis, A.D. and Ouellette, B.F.F., eds., John Wiley & Sons, Inc, 2001; Yona, *et al.*, in *Bioinformatics: Sequence, structure and databanks*, Higgins, D. and Taylor, W. eds, Oxford University Press, 2000; and *Bioinformatics: Sequence and Genome Analysis*, Mount, D.W., ed., Cold Spring Harbor Laboratory Press, 2001).

25 Methods to determine amino acid sequence identity are codified in publicly available computer programs such as GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.), BLAST (Altschul *et al.*, *J. Mol. Biol.* 215(3):403-10, 1990), and FASTA (Pearson, *Methods in Enzymology* 183:63-98, 1990, R.F. Doolittle, ed.).

30 In an embodiment of the present invention sequence identity between two polypeptides is determined using the GAP program (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). GAP uses the alignment method of Needleman and Wunsch. (Needleman, *et al.*, *J. Mol. Biol.* 48:443-453, 1970.) GAP considers all possible alignments and gap positions between two 35 sequences and creates a global alignment that maximizes the number of matched

residues and minimizes the number and size of gaps. A scoring matrix is used to assign values for symbol matches. In addition, a gap creation penalty and a gap extension penalty are required to limit the insertion of gaps into the alignment. Default program parameters for polypeptide comparisons using GAP are the 5 BLOSUM62 (Henikoff *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:10915-10919, 1992) amino acid scoring matrix (MATrix=blosum62.cmp), a gap creation parameter (GAPweight=8) and a gap extension parameter (LENgthweight=2).

More preferred HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptides in addition to being substantially similar to SEQ. ID. NO. 1 across their 10 entire length produce individual NS3, NS4A, NS4B, NS5A and NS5B regions that are substantially similar to the corresponding regions present in SEQ. ID. NO. 1. The corresponding regions in SEQ. ID. NO. 1 are provided as follows: Met-NS3 amino acids 1-632; NS4A amino acids 633-686; NS4B amino acids 687-947; NS5A amino acids 948-1394; and NS5B amino acids 1395-1985.

15 In different embodiments a NS3, NS4A, NS4B, NS5A and/or NS5B region has an amino acid identity to the corresponding region in SEQ. ID. NO. 1 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99%, or 100%; or an amino acid difference of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, or 1-20 amino acids.

20 Amino acid modifications to SEQ. ID. NO. 1 preferably maintain all or most of the T-cell antigen regions. Differences in naturally occurring amino acids are due to different amino acid side chains (R groups). An R group affects different properties of the amino acid such as physical size, charge, and hydrophobicity. Amino acids can be divided into different groups as follows: neutral and hydrophobic 25 (alanine, valine, leucine, isoleucine, proline, tryptophan, phenylalanine, and methionine); neutral and polar (glycine, serine, threonine, tyrosine, cysteine, asparagine, and glutamine); basic (lysine, arginine, and histidine); and acidic (aspartic acid and glutamic acid).

Generally, in substituting different amino acids it is preferable to 30 exchange amino acids having similar properties. Substituting different amino acids within a particular group, such as substituting valine for leucine, arginine for lysine, and asparagine for glutamine are good candidates for not causing a change in polypeptide tertiary structure.

Starting with a particular amino acid sequence and the known 35 degeneracy of the genetic code, a large number of different encoding nucleic acid

- sequences can be obtained. The degeneracy of the genetic code arises because almost all amino acids are encoded by different combinations of nucleotide triplets or "codons". The translation of a particular codon into a particular amino acid is well known in the art (see, e.g., Lewin *GENES IV*, p. 119, Oxford University Press, 1990).
- 5 Amino acids are encoded by codons as follows:
- A=Ala=Alanine: codons GCA, GCC, GCG, GCU
- C=Cys=Cysteine: codons UGC, UGU
- D=Asp=Aspartic acid: codons GAC, GAU
- E=Glu=Glutamic acid: codons GAA, GAG
- 10 F=Phe=Phenylalanine: codons UUC, UUU
- G=Gly=Glycine: codons GGA, GGC, GGG, GGU
- H=His=Histidine: codons CAC, CAU
- I=Ile=Isoleucine: codons AUA, AUC, AUU
- K=Lys=Lysine: codons AAA, AAG
- 15 L=Leu=Leucine: codons UUA, UUG, CUA, CUC, CUG, CUU
- M=Met=Methionine: codon AUG
- N=Asn=Asparagine: codons AAC, AAU
- P=Pro=Proline: codons CCA, CCC, CCG, CCU
- Q=Gln=Glutamine: codons CAA, CAG
- 20 R=Arg=Arginine: codons AGA, AGG, CGA, CGC, CGG, CGU
- S=Ser=Serine: codons AGC, AGU, UCA, UCC, UCG, UCU
- T=Thr=Threonine: codons ACA, ACC, ACG, ACU
- V=Val=Valine: codons GUA, GUC, GUG, GUU
- W=Trp=Tryptophan: codon UGG
- 25 Y=Tyr=Tyrosine: codons UAC, UAU.
- Nucleic acid sequences can be optimized in an effort to enhance expression in a host. Factors to be considered include C:G content, preferred codons, and the avoidance of inhibitory secondary structure. These factors can be combined in different ways in an attempt to obtain nucleic acid sequences having enhanced expression in a particular host. (See, for example, Donnelly *et al.*, International Publication Number WO 97/47358.)
- The ability of a particular sequence to have enhanced expression in a particular host involves some empirical experimentation. Such experimentation involves measuring expression of a prospective nucleic acid sequence and, if needed, altering the sequence.

B. Encoding Nucleotide Sequences

SEQ. ID. NOs. 2 and 3 provide two examples of nucleotide sequences encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B sequence. The coding sequence of SEQ. ID. NO. 2 is similar (99.4% nucleotide sequence identity) to the NS3-NS4A-NS4B-NS5A-NS5B region of the naturally occurring HCV-BK sequence (GenBank accession number M58335). SEQ. ID. NO. 3 is a codon-optimized version of SEQ. ID. NO. 2. SEQ. ID. NOs. 2 and 3 have a nucleotide sequence identity of 78.3%.

Differences between the HCV-BK NS3-NS4A-NS4B-NS5A-NS5B nucleotide (GenBank accession number M58335) and SEQ. ID. NO. 2, include SEQ. ID. NO. 2 having a ribosome binding site, an ATG methionine codon, a region coding for a modified NS5B catalytic domain, a TAAA stop signal and an additional 30 nucleotide differences. The modified catalytic domain codes for a AlaAlaGly (residues 1711-1713) instead of GlyAspAsp to inactivate NS5B.

A nucleotide sequence encoding a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is preferably substantially similar to the SEQ. ID. NO. 2 coding region. In different embodiments, the nucleotide sequence encoding a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide has a nucleotide sequence identify to the SEQ. ID. NO. 2 coding region of at least 65%, at least 75%, at least 85%, at least 95%, at least 99%, or 100%; or differs from SEQ. ID. NO. 2 by 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides.

Nucleotide differences between a sequence coding Met-NS3-NS4A-NS4B-NS5A-NS5B and the SEQ. ID. NO. 2 coding region are calculated by determining the minimum number of nucleotide modifications in which the two sequences differ. Nucleotide modifications can be deletions, additions, substitutions or any combination thereof.

Nucleotide sequence identity is determined by methods well known in the art that compare the nucleotide sequence of one sequence to the nucleotide sequence of a second sequence and generate a sequence alignment. Sequence identity is determined from the alignment by counting the number of aligned positions having identical nucleotides.

Methods for determining nucleotide sequence identity between two polynucleotides include those described by Schuler, in *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Baxevanis, A.D. and Ouellette, B.F.F.,

eds., John Wiley & Sons, Inc, 2001; Yona *et al.*, in *Bioinformatics: Sequence, structure and databanks*, Higgins, D. and Taylor, W. eds, Oxford University Press, 2000; and *Bioinformatics: Sequence and Genome Analysis*, Mount, D.W., ed., Cold Spring Harbor Laboratory Press, 2001). Methods to determine nucleotide sequence identity are codified in publicly available computer programs such as GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.), BLAST (Altschul *et al.*, *J. Mol. Biol.* 215(3):403-10, 1990), and FASTA (Pearson, W.R., *Methods in Enzymology* 183:63-98, 1990, R.F. Doolittle, ed.).

- In an embodiment of the present invention, sequence identity between two polynucleotides is determined by application of GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). GAP uses the alignment method of Needleman and Wunsch. (Needleman *et al.*, *J. Mol. Biol.* 48:443-453, 1970.) GAP considers all possible alignments and gap positions between two sequences and creates a global alignment that maximizes the number of matched residues and minimizes the number and size of gaps. A scoring matrix is used to assign values for symbol matches. In addition, a gap creation penalty and a gap extension penalty are required to limit the insertion of gaps into the alignment. Default program parameters for polynucleotide comparisons using GAP are the nwsgapdna.cmp scoring matrix (MATrix=nwsgapdna.cmp), a gap creation parameter (GAPweight=50) and a gap extension parameter (LENgthweight=3).

More preferred HCV Met-NS3-NS4A-NS4B-NS5A-NS5B nucleotide sequences in addition to being substantially similar across its entire length, produce individual NS3, NS4A, NS4B, NS5A and NS5B regions that are substantially similar to the corresponding regions present in SEQ. ID. NO. 2. The corresponding coding regions in SEQ. ID. NO. 2 are provided as follows: Met-NS3, nucleotides 7-1902; NS4A nucleotides 1903-2064; NS4B nucleotides 2065-2847; NS5A nucleotides 2848-4188; NS5B nucleotides 4189-5661.

In different embodiments a NS3, NS4A, NS4B, NS5A and/or NS5B encoding region has a nucleotide sequence identity to the corresponding region in SEQ. ID. NO. 2 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or a nucleotide difference to SEQ. ID. NO. 2 of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides.

C. Gene Expression Cassettes

A gene expression cassette contains elements needed for polypeptide expression. Reference to "polypeptide" does not provide a size limitation and includes protein. Regulatory elements present in a gene expression cassette generally 5 include: (a) a promoter transcriptionally coupled to a nucleotide sequence encoding the polypeptide, (b) a 5' ribosome binding site functionally coupled to the nucleotide sequence, (c) a terminator joined to the 3' end of the nucleotide sequence, and (d) a 3' polyadenylation signal functionally coupled to the nucleotide sequence. Additional regulatory elements useful for enhancing or regulating gene expression or polypeptide 10 processing may also be present.

Promoters are genetic elements that are recognized by an RNA polymerase and mediate transcription of downstream regions. Preferred promoters are strong promoters that provide for increased levels of transcription. Examples of 15 strong promoters are the immediate early human cytomegalovirus promoter (CMV), and CMV with intron A. (Chapman *et al.*, *Nucl. Acids Res.* 19:3979-3986, 1991.) Additional examples of promoters include naturally occurring promoters such as the EF1 alpha promoter, the murine CMV promoter, Rous sarcoma virus promoter, and 20 SV40 early/late promoters and the β-actin promoter; and artificial promoters such as a synthetic muscle specific promoter and a chimeric muscle-specific/CMV promoter (Li *et al.*, *Nat. Biotechnol.* 17:241-245, 1999, Hagstrom *et al.*, *Blood* 95:2536-2542, 2000).

The ribosome binding site is located at or near the initiation codon. Examples of preferred ribosome binding sites include CCACCAUGG, 25 CCGCCAUGG, and ACCAUGG, where AUG is the initiation codon. (Kozak, *Cell* 44:283-292, 1986). Another example of a ribosome binding site is GCCACCAUGG (SEQ. ID. NO. 12).

The polyadenylation signal is responsible for cleaving the transcribed RNA and the addition of a poly (A) tail to the RNA. The polyadenylation signal in higher eukaryotes contains an AAUAAA sequence about 11-30 nucleotides from the 30 polyadenylation addition site. The AAUAAA sequence is involved in signaling RNA cleavage. (Lewin, *Genes IV*, Oxford University Press, NY, 1990.) The poly (A) tail is important for the mRNA processing.

Polyadenylation signals that can be used as part of a gene expression cassette include the minimal rabbit β -globin polyadenylation signal and the bovine 35 growth hormone polyadenylation (BGH). (Xu *et al.*, *Gene* 272:149-156, 2001, Post *et*

al., U.S. Patent U. S. 5,122,458.) Additional examples include the Synthetic Polyadenylation Signal (SPA) and SV40 polyadenylation signal. The SPA sequence is as follows: AAUAAAAGAUCUUUUUUCAUUAGAUCUGUGUG UGGGUUUUUGUGUG (SEQ. ID. NO. 13).

- 5 Examples of additional regulatory elements useful for enhancing or regulating gene expression or polypeptide processing that may be present include an enhancer, a leader sequence and an operator. An enhancer region increases transcription. Examples of enhancer regions include the CMV enhancer and the SV40 enhancer. (Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, Xu, *et al.*,
10 *Gene* 272:149-156, 2001.) An enhancer region can be associated with a promoter.
A leader sequence is an amino acid region on a polypeptide that directs the polypeptide into the proteasome. Nucleic acid encoding the leader sequence is 5' of a structural gene and is transcribed along the structural gene. An example of a leader sequences is tPA.
- 15 An operator sequence can be used to regulate gene expression. For example, the Tet operator sequence can be used to repress gene expression.

II. THERAPEUTIC VECTORS

- Nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B
20 polypeptide can be introduced into a patient using vectors suitable for therapeutic administration. Suitable vectors can deliver nucleic acid into a target cell without causing an unacceptable side effect.
Cellular expression is achieved using a gene expression cassette encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide. The gene expression cassette contains regulatory elements for producing and processing a sufficient amount of nucleic acid inside a target cell to achieve a beneficial effect.
Examples of vectors that can be used for therapeutic applications include first and second generation adenovectors, helper dependent adenovectors, adeno-associated viral vectors, retroviral vectors, alpha virus vectors, Venezuelan
30 Equine Encephalitis virus vector, and plasmid vectors. (Hitt, *et al.*, *Advances in Pharmacology* 40:137-206, 1997, Johnston *et al.*, U.S. Patent No. 6,156,588, and Johnston *et al.*, International Publication Number WO 95/32733.) Preferred vectors for introducing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide into a subject are first generation adenoviral vectors and plasmid DNA vectors.

A. First Generation Adenovectors

First generation adenovector for expressing a gene expression cassette contain the expression cassette in an E1 and optionally E3 deleted recombinant adenovirus genome. The deletion in the E1 region is sufficiently large to remove 5 elements needed for adenoviral replication.

First generation adenovectors for expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide contain a E1 and E3 deleted recombinant adenovirus genome. The deletion in the E1 region is sufficiently large to remove elements 10 needed for adenoviral replication. The combinations of deletions of the E1 and E3 regions are sufficiently large to accommodate a gene expression cassette encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

The adenovirus has a double-stranded linear genome with inverted terminal repeats at both ends. During viral replication, the genome is packaged inside 15 a viral capsid to form a virion. The virus enters its target cell through viral attachment followed by internalization. (Hitt *et al.*, *Advances in Pharmacology* 40:137-206, 1997.)

Adenovectors can be based on different adenovirus serotypes such as those found in humans or animals. Examples of animal adenoviruses include bovine, porcine, chimp, murine, canine, and avian (CELO). Preferred adenovectors are based 20 on human serotypes, more preferably Group B, C, or D serotypes. Examples of human adenovirus Group B, C, D, or E serotypes include types 2 ("Ad2"), 4 ("Ad4"), 5 ("Ad5"), 6 ("Ad6"), 24 ("Ad24"), 26 ("Ad26"), 34 ("Ad34") and 35 ("Ad35"). Adenovectors can contain regions from a single adenovirus or from two or more adenovirus.

In different embodiments adenovectors are based on Ad5, Ad6, or a combination thereof. Ad5 is described by Chroboczek, *et al.*, *J. Virology* 186:280-285, 1992. Ad6 is described in Figures 7A-7N. An Ad6 based vector containing 25 Ad5 regions is described in the Example section provided below.

Adenovectors do not need to have their E1 and E3 regions completely removed. Rather, a sufficient amount the E1 region is removed to render the vector 30 replication incompetent in the absence of the E1 proteins being supplied in *trans*; and the E1 deletion or the combination of the E1 and E3 deletions are sufficiently large enough to accommodate a gene expression cassette.

E1 deletions can be obtained starting at about base pair 342 going up to 35 about base pair 3523 of Ad5, or a corresponding region from other adenoviruses.

Preferably, the deleted region involves removing a region from about base pair 450 to about base pair 3511 of Ad5, or a corresponding region from other adenoviruses.

Larger E1 region deletions starting at about base pair 341 removes elements that facilitate virus packaging.

5 E3 deletions can be obtained starting at about base pair 27865 to about base pair 30995 of Ad5, or the corresponding region of other adenovectors.

Preferably the deletion region involves removing a region from about base pair 28134 up to about base pair 30817 of Ad5, or the corresponding region of other adenovectors.

10 The combination of deletions to the E1 region and optionally the E3 region should be sufficiently large so that the overall size of the recombinant genome containing the gene expression cassette does not exceed about 105% of the wild type adenovirus genome. For example, as recombinant adenovirus Ad5 genomes increase size above about 105% the genome becomes unstable. (Bett *et al.*, *Journal of*

15 *Virology* 67:5911-5921, 1993.)

Preferably, the size of the recombinant adenovirus genome containing the gene expression cassette is about 85% to about 105% the size of the wild type adenovirus genome. In different embodiments, the size of the recombinant adenovirus genome containing the expression cassette is about 100% to about

20 105.2%, or about 100%, the size of the wild type genome.

Approximately 7,500 kb can be inserted into an adenovirus genome with a E1 and E3 deletion. Without any deletion, the Ad5 genome is 35,935 base pairs and the Ad6 genome is 35,759 base pairs.

25 Replication of first generation adenovectors can be performed by supplying the E1 gene products in *trans*. The E1 gene product can be supplied in *trans*, for example, by using cell lines that have been transformed with the adenovirus E1 region. Examples of cells and cells lines transformed with the adenovirus E1 region are HEK 293 cells, 911 cells, PERC.6™ cells, and transfected primary human aminocytes cells. (Graham *et al.*, *Journal of Virology* 36:59-72, 1977, Schiedner *et*

30 *al.*, *Human Gene Therapy* 11:2105-2116, 2000, Fallaux *et al.*, *Human Gene Therapy* 9:1909-1917, 1998, Bout *et al.*, U.S. Patent No. 6,033,908.)

A Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette should be inserted into a recombinant adenovirus genome in the region corresponding to the deleted E1 region or the deleted E3 region. The expression cassette can have a parallel or anti-parallel orientation. In a parallel orientation the transcription direction

of the inserted gene is the same direction as the deleted E1 or E3 gene. In an anti-parallel orientation transcription the opposite strand serves as a template and the transcription direction is in the opposite direction.

- In an embodiment of the present invention the adenovector has a gene expression cassette inserted in the E1 deleted region. The vector contains:
- 5 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
 - 10 c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - 15 e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region; and
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6 joined to the fourth region.

In another embodiment of the present invention the adenovector has an expression cassette inserted in the E3 deleted region. The vector contains:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- 30 d) a gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;

- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about 5 base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region.

In preferred different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first 10 region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

B. DNA Plasmid Vectors

15 DNA vaccine plasmid vectors contain a gene expression cassette along with elements facilitating replication and preferably vector selection. Preferred elements provide for replication in non-mammalian cells and a selectable marker. The vectors should not contain elements providing for replication in human cells or for integration into human nucleic acid.

20 The selectable marker facilitates selection of nucleic acids containing the marker. Preferred selectable markers are those that confer antibiotic resistance. Examples of antibiotic selection genes include nucleic acid encoding resistance to ampicillin, neomycin, and kanamycin.

Suitable DNA vaccine vectors can be produced starting with a plasmid 25 containing a bacterial origin of replication and a selectable marker. Examples of bacterial origins of replication providing for higher yields include the ColE1 plasmid-derived bacterial origin of replication. (Donnelly *et al.*, *Annu. Rev. Immunol.* 15:617-648, 1997.)

The presence of the bacterial origin of replication and selectable 30 marker allows for the production of the DNA vector in a bacterial strain such as *E. coli*. The selectable marker is used to eliminate bacteria not containing the DNA vector.

III. AD6 RECOMBINANT NUCLEIC ACID

Ad6 recombinant nucleic acid comprises an Ad6 region substantially similar to an Ad6 region found in SEQ. ID. NO. 8, and a region not present in Ad6 nucleic acid. Recombinant nucleic acid comprising Ad6 regions have different uses
5 such as in producing different Ad6 regions, as intermediates in the production of Ad6 based vectors, and as a vector for delivering a recombinant gene.

As depicted in Figure 9, the genomic organization of Ad6 is very similar to the genomic organization of Ad5. The homology between Ad5 and Ad6 is approximately 98%.

10 In different embodiments, the Ad6 recombinant nucleic acid comprises a nucleotide region substantially similar to E1A, E1B, E2B, E2A, E3, E4, L1, L2, L3, or L4, or any combination thereof. A substantially similar nucleic acid region to an Ad6 region has a nucleotide sequence identity of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or a nucleotide difference of 1-2, 1-3, 1-4,
15 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides. Techniques and embodiments for determining substantially similar nucleic acid sequences are described in Section I.B.
supra.

20 Preferably, the recombinant Ad6 nucleic acid contains an expression cassette coding for a polypeptide not found in Ad6. Examples of expression cassettes include those coding for HCV regions and those coding for other types of polypeptides.

25 Different types of adenoviral vectors can be produced incorporating different amounts of Ad6, such as first and second generation adenovectors. As noted in Section II.A. *supra.* first generation adenovectors are defective in E1 and can replicate when E1 is supplied *in trans.*

30 Second generation adenovectors contain less adenoviral genome than first generation vectors and can be used in conjugation with complementing cell lines and/or helper vectors supplying adenoviral proteins. Second generation adenovectors are described in different references such as Russell, *Journal of General Virology* 81:2573-2604, 2000; Hitt *et al.*, 1997, Human Ad vectors for Gene Transfer, *Advances in Pharmacology*, Vol 40 Academic Press.

In an embodiment of the present invention, the Ad6 recombinant nucleic acid is an adenovirus vector defective in E1 that is able to replicate when E1 is

supplied *in trans*. Expression cassettes can be inserted into a deleted E1 region and/or a deleted E3 region.

An example of an Ad6 based adenoviral vector with an expression cassette provided in a deleted E1 region comprises or consists of:

- 5 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
 - c) a second adenovirus region from about base pair 3511 to about
 - 10 base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - 15 e) an optionally present fourth region from about base pair 28134 to about base pair 30817 corresponding to Ad5, or from about base pair 28157 to about base pair 30788 corresponding to Ad6, joined to the third region;
 - f) a fifth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base
 - 20 pair 33784 corresponding to Ad6, wherein the fifth region is joined to the fourth region if the fourth region is present, or the fifth is joined to the third region if the fourth region is not present; and
 - g) a sixth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base
 - 25 pair 35759 corresponding to Ad6, joined to the fifth region;
- wherein at least one Ad6 region is present.

In different embodiments of the invention, all of the regions are from Ad6; all of the regions except for the first and second are from Ad6; and 1, 2, 3, or 4 regions selected from the second, third, fourth, and fifth regions are from Ad6.

- 30 An example of an Ad6 based adenoviral vector with an expression cassette provided in a deleted E3 region comprises or consists of:
 - a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the first region;
- c) a third adenovirus region from about base pair 5549 to about 5 base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- d) a gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;
- e) a fourth adenovirus region from about base pair 30818 to about 10 base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region;
- 15 wherein at least one Ad6 region is present.
- In different embodiment of the invention, all of the regions are from Ad6; all of the regions expect for the first and second are from Ad6; and 1, 2, 3, or 4 regions selected from the second, third, fourth and fifth regions are from Ad6.

20

IV. VECTOR PRODUCTION

Vectors can be produced using recombinant nucleic acid techniques such as those involving the use of restriction enzymes, nucleic acid ligation, and homologous recombination. Recombinant nucleic acid techniques are well known in the art. (Ausubel, *Current Protocols in Molecular Biology*, John Wiley, 1987-1998, and Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, 1989.)

25
30

Intermediate vectors are used to derive a therapeutic vector or to transfer an expression cassette or portion thereof from one vector to another vector. Examples of intermediate vectors include adenovirus genome plasmids and shuttle vectors.

Useful elements in an intermediate vector include an origin of replication, a selectable marker, homologous recombination regions, and convenient restriction sites. Convenient restriction sites can be used to facilitate cloning or release of a nucleic acid sequence.

Homologous recombination regions provide nucleic acid sequence regions that are homologous to a target region in another nucleic acid molecule. The homologous regions flank the nucleic acid sequence that is being inserted into the target region. In different embodiments homologous regions are preferably about 150 to 600 nucleotides in length, or about 100 to 500 nucleotides in length.

An embodiment of the present invention describes a shuttle vector containing a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette, a selectable marker, a bacterial origin of replication, a first adenovirus homology region and a second adenovirus homologous region that target the expression cassette to insert in or replace an E1 region. The first and second homology regions flank the expression cassette. The first homology region contains at least about 100 base pairs substantially homologous to at least the right end (3' end) of a wild-type adenovirus region from about base pairs 4-450. The second homology contains at least about 100 base pairs substantially homologous to at least the left end (5' end) of Ad5 from about base pairs 3511-5792, or the corresponding region from another adenovirus.

Reference to "substantially homologous" indicates a sufficient degree of homology to specifically recombine with a target region. In different embodiments substantially homologous refers to at least 85%, at least 95%, or 100% sequence identity. Sequence identity can be calculated as described in Section I.B. *supra*.

One method of producing adenovectors is through the creation of an adenovirus genome plasmid containing an expression cassette. The pre-Adenovirus plasmid contains all the adenovirus sequences needed for replication in the desired complimenting cell line. The pre-Adenovirus plasmid is then digested with a restriction enzyme to release the viral ITR's and transfected into the complementing cell line for virus rescue. The ITR's must be released from plasmid sequences to allow replication to occur. Adenovector rescue results in the production on an adenovector containing the expression cassette.

A. Adenovirus Genome Plasmids

Adenovirus genome plasmids contain an adenovector sequence inside a longer-length plasmid (which may be a cosmid). The longer-length plasmid may contain additional elements such as those facilitating growth and selection in eukaryotic or bacterial cells depending upon the procedures employed to produce and maintain the plasmid. Techniques for producing adenovirus genome plasmids include those involving the use of shuttle vectors and homologous recombination, and those

involving the insertion of a gene expression cassette into an adenovirus cosmid. (Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, Danthinne *et al.*, *Gene Therapy* 7:1707-1714, 2000.)

- Adenovirus genome plasmids preferably have a gene expression cassette inserted into a E1 or E3 deleted region. In an embodiment of the present invention, the adenovirus genome plasmid contains a gene expression cassette inserted in the E1 deleted region, an origin of replication, a selectable marker, and the recombinant adenovirus region is made up of:
- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
 - c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region;
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region, and
 - g) an optionally present E3 region corresponding to all or part of the E3 region present in Ad5 or Ad6, which may be present for smaller inserts taking into account the overall size of the desired adenovector.

In another embodiment of the present invention the recombinant adenovirus genome plasmid has the gene expression cassette inserted in the E3 deleted region. The vector contains an origin of replication, a selectable marker, and the following:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
 - c) a third adenovirus region from about base pair 5549 to about 5 base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - d) the gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;
 - e) a fourth adenovirus region from about base pair 30818 to about 10 base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region.
- 15 In different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region 20 corresponds to Ad5.

An embodiment of the present invention describes a method of making an adenovector involving a homologous recombination step to produce a adenovirus genome plasmid and an adenovirus rescue step. The homologous recombination step involves the use of a shuttle vector containing a Met-NS3-NS4A-NS4B-NS5A-NS5B 25 expression cassette flanked by adenovirus homology regions. The adenovirus homology regions target the expression cassette into either the E1 or E3 deleted region.

In an embodiment of the present invention concerning the production of an adenovirus genome plasmid, the gene expression cassette is inserted into a 30 vector comprising: a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6; a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the second region; a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding 35 to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6,

- joined to the second region; a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region; and a fifth adenovirus region from about 33967 to about 35935 corresponding to Ad5 or from 5 about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region. The adenovirus genome plasmid should contain an origin of replication and a selectable marker, and may contain all or part of the Ad5 or Ad6 E3 region.

In different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the 10 first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

15 **B. Adenovector Rescue**

An adenovector can be rescued from a recombinant adenovirus genome plasmid using techniques known in the art or described herein. Examples of techniques for adenovirus rescue well known in the art are provided by Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, and Danthinne *et al.*, *Gene Therapy* 20 7:1707-1714, 2000.

A preferred method of rescuing an adenovector described herein involves boosting adenoviral replication. Boosting adenoviral replication can be performed, for example, by supplying adenoviral functions such as E2 proteins (polymerase, pre-terminal protein and DNA binding protein) as well as E4 orf6 on a 25 separate plasmid. Example 10 *infra*. illustrates the boosting of adenoviral replication to rescue an adenovector containing a codon optimized Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette.

V. PARTIAL-OPTIMIZED HCV ENCODING SEQUENCES

30 Partial optimization of HCV polyprotein encoding nucleic acid provides for a lesser amount of codons optimized for expression in a human than complete optimization. The overall objective is to provide the benefits of increased expression due to codon optimization, while facilitating the production of an adenovector containing HCV polyprotein encoding nucleic acid having optimized 35 codons.

Complete optimization of an HCV polyprotein encoding sequence provides the most frequently observed human codon for each amino acid. Complete optimization can be performed using codon frequency tables well known in the art and using programs such as the BACKTRANSLATE program (Wisconsin Package 5 version 10, Genetics Computer Group, GCG, Madison, Wisc.).

Partial optimization can be preformed on an entire HCV polyprotein encoding sequence that is present (e.g., NS3-NS5B), or one or more local regions that are present. In different embodiments the GC content for the entire HCV encoded polyprotein that is present is no greater than at least about 65%; and the GC content for one or more 10 local regions is no greater than about 70%.

Local regions are regions present in HCV encoding nucleic acid, and can vary in size. For example, local regions can be about 60, about 70, about 80, about 90 or about 100 nucleotides in length.

Partial optimization can be achieved by initially constructing an HCV 15 encoding polyprotein sequence to be partially optimized based on a naturally occurring sequence. Alternatively, an optimized HCV encoding sequence can be used as basis of comparison to produce a partial optimized sequence.

VI. HCV COMBINATION TREATMENT

The HCV Met-NS3-NS4A-NS4B-NS5A-NS5B vaccine can be used by 20 itself to treat a patient, can be used in conjunction with other HCV therapeutics, and can be used with agents targeting other types of diseases. Additional therapeutics include additional therapeutic agents to treat HCV and diseases having a high prevalence in HCV infected persons. Agents targeting other types of disease include 25 vaccines directed against HIV and HBV.

Additional therapeutics for treating HCV include vaccines and non-vaccine agents. (Zein, *Expert Opin. Investig. Drugs* 10:1457-1469, 2001.) Examples of additional HCV vaccines include vaccines designed to elicit an immune response against an HCV core antigen and the HCV E1, E2 or p7 region. Vaccine components 30 can be naturally occurring HCV polypeptides, HCV mimotope polypeptides or nucleic acid encoding such polypeptides.

HCV mimotope polypeptides contain HCV epitopes, but have a different sequence than a naturally occurring HCV antigen. A HCV mimotope can be fused to a naturally occurring HCV antigen. References describing techniques for 35 producing mimotopes in general and describing different HCV mimotopes are

provided in Felici *et al.* U.S. Patent No. 5,994,083 and Nicosia *et al.*, International Application Number WO 99/60132.

VII. PHARMACEUTICAL ADMINISTRATION

5 HCV vaccines can be formulated and administered to a patient using the guidance provided herein along with techniques well known in the art. Guidelines for pharmaceutical administration in general are provided in, for example, *Modern Vaccinology*, Ed. Kurstak, Plenum Med. Co. 1994; *Remington's Pharmaceutical Sciences 18th Edition*, Ed. Gennaro, Mack Publishing, 1990; and *Modern Pharmaceutics 2nd Edition*, Eds. Bunker and Rhodes, Marcel Dekker, Inc., 1990, each of which are hereby incorporated by reference herein.

10 HCV vaccines can be administered by different routes such intravenous, intraperitoneal, subcutaneous, intramuscular, intradermal, impression through the skin, or nasal. A preferred route is intramuscular.

15 Intramuscular administration can be performed using different techniques such as by injection with or without one or more electric pulses. Electric mediated transfer can assist genetic immunization by stimulating both humoral and cellular immune responses.

20 Vaccine injection can be performed using different techniques, such as by employing a needle or a needless injection system. An example of a needless injection system is a jet injection device. (Donnelly *et al.*, International Publication Number WO 99/52463.)

A. Electrically Mediated Transfer

25 Electrically mediated transfer or Gene Electro-Transfer (GET) can be performed by delivering suitable electric pulses after nucleic acid injection. (See Mathiesen, International Publication Number WO 98/43702). Plasmid injection and electroporation can be performed using stainless needles. Needles can be used in couples, triplets or more complex patterns. In one configuration the needles are 30 soldered on a printed circuit board that is a mechanical support and connects the needles to the electrical field generator by means of suitable cables.

The electrical stimulus is given in the form of electrical pulses. Pulses can be of different forms (square, sinusoidal, triangular, exponential decay) and different polarity (monopolar or positive or negative polarity, bipolar). Pulses can be 35 delivered either at constant voltage or constant current modality.

Different patterns of electric treatment can be used to introduce nucleic acid vaccines including HCV and other nucleic acid vaccines into a patient. Possible patterns of electric treatment include the following:

- 5 Treatment 1: 10 trains of 1000 square bipolar pulses delivered every other second, pulse length 0.2 msec/phase, frequency 1000 Hz, constant voltage mode, 45 Volts/phase, floating current.
- 10 Treatment 2: 2 trains of 100 square bipolar pulses delivered every other second, pulse length 2 msec/phase, frequency 100 Hz, constant current mode, 100 mA/phase, floating voltage.
- 15 Treatment 3: 2 trains of bipolar pulses at a pulse length of about 2 msec/phase, for a total length of about 3 seconds, where the actual current going through the tissue is fixed at about 50 mA.
- 20 Electric pulses are delivered through an electric field generator. A suitable generator can be composed of three independent hardware elements assembled in a common chassis and driven by a portable PC which runs the driving program. The software manages both basic and accessory functions. The elements of the device are: (1) signal generator driven by a microprocessor, (2) power amplifier and (3) digital oscilloscope.
- 25 The signal generator delivers signals having arbitrary frequency and shape in a given range under software control. The same software has an interactive editor for the waveform to be delivered. The generator features a digitally controlled current limiting device (a safety feature to control the maximal current output). The power amplifier can amplify the signal generated up to +/- 150 V. The oscilloscope is digital and is able to sample both the voltage and the current being delivered by the amplifier.

B. Pharmaceutical Carriers

- 30 Pharmaceutically acceptable carriers facilitate storage and administration of a vaccine to a subject. Examples of pharmaceutically acceptable carriers are described herein. Additional pharmaceutical acceptable carriers are well known in the art.

- 35 Pharmaceutically acceptable carriers may contain different components such a buffer, normal saline or phosphate buffered saline, sucrose, salts and polysorbate. An example of a pharmaceutically acceptable carrier is follows: 2.5-10 mM TRIS buffer, preferably about 5 mM TRIS buffer; 25-100 mM NaCl, preferably

about 75 mM NaCl; 2.5-10% sucrose, preferably about 5% sucrose; 0.01 -2 mM MgCl₂; and 0.001%-0.01% polysorbate 80 (plant derived). The pH is preferably from about 7.0-9.0, more preferably about 8.0. A specific example of a carrier contains 5 mM TRIS, 75 mM NaCl, 5% sucrose, 1 mM MgCl₂, 0.005% polysorbate 80 at pH 5 8.0.

C. Dosing Regimes

Suitable dosing regimens can be determined taking into account the efficacy of a particular vaccine and factors such as age, weight, sex and medical 10 condition of a patient; the route of administration; the desired effect; and the number of doses. The efficacy of a particular vaccine depends on different factors such as the ability of a particular vaccine to produce polypeptide that is expressed and processed in a cell and presented in the context of MHC class I and II complexes.

HCV encoding nucleic acid administered to a patient can be part of 15 different types of vectors including viral vectors such as adenovector, and DNA plasmid vaccines. In different embodiments concerning administration of a DNA plasmid, about 0.1 to 10 mg of plasmid is administered to a patient, and about 1 to 5 mg of plasmid is administered to a patient. In different embodiments concerning administration of a viral vector, preferably an adenoviral vector, about 10⁵ to 10¹¹ 20 viral particles are administered to a patient, and about 10⁷ to 10¹⁰ viral particles are administered to a patient.

Viral vector vaccines and DNA plasmid vaccines may be administered alone, or may be part of a prime and boost administration regimen. A mixed modality 25 priming and booster inoculation involves either priming with a DNA vaccine and boosting with viral vector vaccine, or priming with a viral vector vaccine and boosting with a DNA vaccine.

Multiple priming, for example, about to 2-4 or more may be used. The length of time between priming and boost may typically vary from about four months to a year, but other time frames may be used. The use of a priming regimen with a 30 DNA vaccine may be preferred in situations where a person has a pre-existing antiadenovirus immune response.

In an embodiment of the present invention, 1x10⁷ to 1x10¹² particles and preferably about 1x10¹⁰ to 1x10¹¹ particles of adenovector is administered directly 35 into muscle tissue. Following initial vaccination a boost is performed with an adenovector or DNA vaccine.

In another embodiment of the present invention initial vaccination is performed with a DNA vaccine directly into muscle tissue. Following initial vaccination a boost is performed with an adenovector or DNA vaccine.

- Agents such as interleukin-12, GM-CSF, B7-1, B7-2, IP10, Mig-1 can 5 be coadministered to boost the immune response. The agents can be coadministered as proteins or through use of nucleic acid vectors.

D. Heterologous Prime-Boost

Heterologous prime-boost is a mixed modality involving the use of one 10 type of viral vector for priming and another type of viral vector for boosting. The heterologous prime-boost can involve related vectors such as vectors based on different adenovirus serotypes and more distantly related viruses such adenovirus and poxvirus. The use of poxvirus and adenovirus vectors to protect mice against malaria is illustrated by Gilbert *et al.*, *Vaccine* 20:1039-1045, 2002.

15 Different embodiments concerning priming and boosting involve the following types of vectors expressing desired antigens such as Met-NS3-NS4A-NS4B-NS5A-NS5B: Ad5 vector followed by Ad6 vector; Ad6 vector followed by Ad5 vector; Ad5 vector followed by poxvirus vector; poxvirus vector followed by Ad5 vector; Ad6 vector followed by poxvirus vector; and poxvirus vector followed by 20 Ad6 vector.

The length of time between priming and boosting typically varies from about four months to a year, but other time frames may be used. The minimum time frame should be sufficient to allow for an immunological rest. In an embodiment, this rest is for a period of at least 6 months. Priming may involve multiple priming with 25 one type of vector, such as 2-4 primings.

Expression cassettes present in a poxvirus vector should contain a promoter either native to, or derived from, the poxvirus of interest or another poxvirus member. Different strategies for constructing and employing different types of poxvirus based vectors including those based on vaccinia virus, modified vaccinia 30 virus, avipoxvirus, raccoon poxvirus, modified vaccinia virus Ankara, canarypoxviruses (such as ALVAC), fowlpoxviruses, cowpoxviruses, and NYVAC are well known in the art. (Moss, *Current Topics in Microbiology and Immunology* 158:25-38, 1982; Earl *et al.*, In *Current Protocols in Molecular Biology*, Ausubel *et al.* eds., New York: Greene Publishing Associates & Wiley Interscience; 35 1991:16.16.1-16.16.7; Child *et al.*, *Virology* 174(2):625-9, 1990; Tartaglia *et al.*,

Virology 188:217-232, 1992; U.S. Patent Nos., 4,603,112, 4,722,848, 4,769,330, 5,110,587, 5,174,993, 5,185,146, 5,266,313, 5,505,941, 5,863,542, and 5,942,235.

E. Adjuvants

- 5 HCV vaccines can be formulated with an adjuvant. Adjuvants are particularly useful for DNA plasmid vaccines. Examples of adjuvants are alum, AlPO₄, alhydrogel, Lipid-A and derivatives or variants thereof, Freund's incomplete adjuvant, neutral liposomes, liposomes containing the vaccine and cytokines, non-ionic block copolymers, and chemokines.
- 10 Non-ionic block polymers containing polyoxyethylene (POE) and polyxylpropylene (POP), such as POE-POP-POE block copolymers may be used as an adjuvant. (Newman *et al.*, *Critical Reviews in Therapeutic Drug Carrier Systems* 15:89-142, 1998.) The immune response of a nucleic acid can be enhanced using a non-ionic block copolymer combined with an anionic surfactant.
- 15 A specific example of an adjuvant formulation is one containing CRL-1005 (CytRx Research Laboratories), DNA, and benzylalkonium chloride (BAK). The formulation can be prepared by adding pure polymer to a cold (< 5°C) solution of plasmid DNA in PBS using a positive displacement pipette. The solution is then vortexed to solubilize the polymer. After complete solubilization of the polymer a clear solution is obtained at temperatures below the cloud point of the polymer (~6-7°C). Approximately 4 mM BAK is then added to the DNA/CRL-1005 solution in PBS, by slow addition of a dilute solution of BAK dissolved in PBS. The initial DNA concentration is approximately 6 mg/mL before the addition of polymer and BAK, and the final DNA concentration is about 5 mg/mL. After BAK addition the formulation is vortexed extensively, while the temperature is allowed to increase from ~ 2°C to above the cloud point. The formulation is then placed on ice to decrease the temperature below the cloud point. Then, the formulation is vortexed while the temperature is allowed to increase from ~2°C to above the cloud point. Cooling and mixing while the temperature is allowed to increase from ~2°C to above the cloud point is repeated several times, until the particle size of the formulation is about 200-500 nm, as measured by dynamic light scattering. The formulation is then stored on ice until the solution is clear, then placed in storage at -70°C. Before use, the formulation is allowed to thaw at room temperature.

F. Vaccine Storage

Adenovector and DNA vaccines can be stored using different types of buffers. For example, buffer A105 described in Example 9 *infra.* can be used to for vector storage.

- 5 Storage of DNA can be enhanced by removal or chelation of trace metal ions. Reagents such as succinic or malic acid, and chelators can be used to enhance DNA vaccine stability. Examples of chelators include multiple phosphate ligands and EDTA. The inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, can also be useful to prevent damage of DNA plasmid from free
10 radical production. Furthermore, the buffer type, pH, salt concentration, light exposure, as well as the type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine.

VII. EXAMPLES

- 15 Examples are provided below to further illustrate different features of the present invention. The examples also illustrate useful methodology for practicing the invention. These examples do not limit the claimed invention.

Example 1: Met-NS3-NS4A-NS4B-NS5A-NS5B Expression Cassettes

- 20 Different gene expression cassettes encoding HCV NS3-NS4A-NS4B-NS5A-NS5B were constructed based on a 1b subtype HCV BK strain. The encoded sequences had either (1) an active NS5B sequence ("NS"), (2) an inactive NS5B sequence ("NSmut"), (3) a codon optimized sequence with an inactive NS5B sequence ("NSOPTmut"). The expression cassettes also contained a CMV promoter/enhancer and the BGH polyadenylation signal.

25 The NS nucleotide sequence (SEQ. ID. NO. 5) differs from HCV BK strain GenBank accession number M58335 by 30 out of 5952 nucleotides. The NS amino acid sequence (SEQ. ID. NO. 6) differs from the corresponding 1b genotype HCV BK strain by 7 out of 1984 amino acids. To allow for initiation of translation an ATG codon is present at the 5' end of the NS sequence. A TGA termination sequence is present at the 3' end of the NS sequence.

30 The NSmut nucleotide sequence (SEQ. ID. NO. 2, Figure 2), is similar to the NS sequence. The differences between NSmut and NS include NSmut having an altered NS5B catalytic site; an optimal ribosome binding site at the 5' end; and a TAAA termination sequence at the 3' end. The alterations in NS5B comprise bases

5138 to 5146, which encode amino acids 1711 to 1713. The alterations result in a change of amino acids GlyAspAsp into AlaAlaGly and creates an inactive form of the NS5B RNA-dependent RNA-polymerase NS5B.

The NSOPTmut sequence (SEQ. ID. NO. 3, Figure 3) was designed
5 based on the amino acid sequence encoded by NSmut. The NSmut amino acid sequence was back translated into a nucleotide sequence with the GCG (Wisconsin Package version 10, Genetics Computer Group, GCG, Madison, Wisc.) BACKTRANSLATE program. To generate a NSOPTmut nucleotide sequence where each amino acid is coded for by the corresponding most frequently observed human
10 codon, the program was run choosing as parameter the generation of the most probable nucleotide sequence and specifying the codon frequency table of highly expressed human genes (human_high.cod) available within the GCG Package as translation scheme.

15 Example 2: Generation pV1Jns plasmid with NS, NSmut or NSOPTmut Sequences
pV1Jns plasmids containing either the NS sequence, NSmut sequence or NSOPTmut sequences were generated and characterised as follows:

pV1Jns Plasmid with the NS Sequence

20 The coding region Met-NS3-NS4A-NS4B-NS5A and the coding region Met-NS3-NS4A-NS4B-NS5A-NS5B from a HCV BK type strain (Tomei *et al.*, *J. Virol.* 67:4017-4026, 1993) were cloned into pcDNA3 plasmid (Invitrogen), generating pcD3-5a and pcD3-5b vectors, respectively. Pcd3-5A was digested with Hind III, blunt-ended with Klenow fill-in and subsequently digested with Xba I, to generate a fragment corresponding to the coding region of Met-NS3-NS4A-NS4B-NS5A. The fragment was cloned into pV1Jns-poly, digested with Bgl II blunt-ended with Klenow fill-in and subsequently digested with Xba I, generating pV1JnsNS3-5A.

25 pV1Jns-poly is a derivative of pV1JnsA plasmid (Montgomery *et al.*, *DNA and Cell Biol.* 12:777-783, 1993), modified by insertion of a polylinker containing recognition sites for XbaI, PmeI, PacI into the unique BglII and NotI restriction sites. The pV1Jns plasmid with the NS sequence (pV1JnsNS3-5B) was obtained by homologous recombination into the bacterial strain BJ5183, co-transforming pV1JNS3-5A linearized with XbaI and NotI digestion and a PCR fragment containing approximately 200 bp of NS5A, NS5B coding sequence and

approximately 60 bp of the BGH polyadenylation signal. The resulting plasmid represents pV1Jns-NS.

pV1Jns-NS can be summarized as follows:

Bases	1 to 1881 of pV1JnsA
5 an additional	AGCTT
then the	Met-NS3-NS5B sequence (SEQ. ID. NO. 5)
then the	wt TGA stop
an additional	TCTAGAGCGTTAAACCCCTTAATTAAGG (SEQ. ID.
NO. 14)	
10 Bases	1912 to 4909 of pV1JnsA

pV1Jns Plasmid with the NSmut Sequence

The V1JnsNS3-5A plasmid was modified at the 5' of the NS3 coding sequence by addition of a full Kozak sequence. The plasmid (V1JNS3-5Akozak) was obtained by homologous recombination into the bacterial strain BJ5183, co-transforming V1JNS3-5A linearized by *Af*II digestion and a PCR fragment containing the proximal part of Intron A, the restriction site *Bgl*II, a full Kozak translation initiation sequence and part of the NS3 coding sequence.

The resulting plasmid (V1JNS3-5Akozak) was linearized with *Xba* I digestion and co-transformed into the bacterial strain BJ5183 with a PCR fragment, containing approximately 200 bp of NS5A, the NS5B mutated sequence, the strong translation termination TAAA and approximately 60 bp of the BGH polyadenylation signal. The PCR fragment was obtained by assembling two 22bp-overlapping fragments where mutations were introduced by the oligonucleotides used for their amplification. The resulting plasmid represents pV1Jns-NSmut.

pV1Jns-NSmut can be summarized as follows:

Bases	1 to 1882 of pV1JnsA
then the	kozak Met-NS3-NS5B(mut) TAAA sequence (SEQ. ID. NO. 2)
an additional	TCTAGA
30 Bases	1925 to 4909 of pV1JnsA

pV1Jns Plasmid with the NSOPTmut Sequence

The human codon-optimized synthetic gene (NSOPTmut) with mutated NS5B to abrogate enzymatic activity, full Kozak translation initiation sequence and a strong translation termination was digested with *Bam*HII and *Sall*

restriction sites present at the 5' and 3' end of the gene. The gene was then cloned into the BglII and SalI restriction sites present in the polylinker of pV1JnsA plasmid, generating pV1Jns-NSOPTmut.

pV1Jns-NSOPTmut can be summarized as follows:

5 Bases 1 to 1881 of pV1JnsA
an additional C
then kozak Met-NS3-NS5B(optmut) TAAA sequence (SEQ. ID. NO. 3)
an additional TTTAAATGTTAAC (SEQ. ID. NO. 15)
Bases 1905 to 4909 of pV1JnsA

10

Plasmids Characterization

Expression of HCV NS proteins was tested by transfection of HEK 293 cells, grown in 10% FCS/DMEM supplemented by L-glutamine (final 4 mM). Twenty-four hours before transfection, cells were plated in 6-well 35 mm diameter, to reach 90-95% confluence on the day of transfection. Forty nanograms of plasmid DNA (previously assessed as a non-saturating DNA amount) were co-transfected with 100 ng of pRSV-Luc plasmid containing the luciferase reporter gene under the control of Rous sarcoma virus promoter, using the LIPOFECTAMINE 2000 reagent. Cells were kept in a CO₂ incubator for 48 hours at 37 °C.

20 Cell extracts were prepared in 1% Triton/TEN buffer. The extracts were normalized for Luciferase activity, and run in serial dilution on 10% SDS-acrylamide gel. Proteins were transferred on nitrocellulose and assayed with antibodies directed against NS3, NS5A and NS5B to assess strength of expression and correct proteolytic cleavage. Mock-transfected cells were used as a negative control.
25 Results from representative experiments testing pV1JnsNS, pV1JnsNSmut and pV1JnsNSOPTmut are shown in Figure 12.

Example 3: Mice Immunization with Plasmid DNA Vectors

The DNA plasmids pV1Jns-NS, pV1Jns-NSmut and pV1Jns-NSOPTmut were injected in different mice strains to evaluate their potential to elicit anti-HCV immune responses. Two different strains (Balb/C and C57Black6, N=9-10) were injected intramuscularly with 25 or 50 µg of DNA followed by electrical pulses. Each animal received two doses at three weeks interval.

Humoral immune response elicited in C57Black6 mice against the NS3 protein was measured in post dose two sera by ELISA on bacterially expressed NS3

protease domain. Antibodies specific for the tested antigen were detected in animals immunized with all three vectors with geometric mean titers (GMT) ranging from 94000 to 133000 (Tables 1-3).

5

Table 1: pV1jns-NS

Mice n.	1	2	3	4	5	6	7	8	9	GMT
Titer	105466	891980	78799	39496	543542	182139	32351	95028	67800	94553

Table 2: pV1jns-NSmut

10

Mice n.	11	12	13	14	15	16	17	18	19	20	GMT
Titer	202981	55670	130786	49748	17672	174958	44304	37337	78182	193695	75083

Table 3: pV1jns-NSOPTmut

Mice n.	21	22	23	24	25	26	27	28	29	30	GMT
Titer	310349	43645	63496	82174	630778	297259	66861	146735	173506	77732	133165

15

- A T cell response was measured in C57Black6 mice immunized with two intramuscular injections at three weeks interval with 25 µg of plasmid DNA. Quantitative ELispot assay was performed to determine the number of IFN γ secreting T cells in response to five pools of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence. Specific CD8+ response was analyzed by the same assay using a 20mer peptide encompassing a CD8+ epitope for C57Black6 mice (pep1480).

Cells secreting IFN γ in an antigen specific-manner were detected using a standard ELispot assay. T cell response in C57Black6 mice immunized with two intramuscular injections at three weeks interval with 50 µg of plasmid DNA, was

analyzed by the same ELIspot assay measuring the number of IFN γ secreting T cells in response to five pools of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence.

Spleen cells were prepared from immunized mice and re-suspended in
5 R10 medium (RPMI 1640 supplemented with 10% FCS, 2 mM L-Glutamine, 50 U/ml-50 μ g/ml Penicillin/Streptomycin, 10 mM Hepes, 50 μ M 2-mercapto-ethanol). Multiscreen 96-well Filtration Plates (Millipore, Cat. No. MAIPS4510, Millipore Corporation, 80 Ashby Road Bedford, MA) were coated with purified rat anti-mouse IFN γ antibody (PharMingen, Cat. No. 18181D, PharmiMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA). After overnight incubation, plates were washed with PBS 1X/0.005% Tween and blocked with 250 μ l/well of R10 medium.
10

Splenocytes from immunized mice were prepared and incubated for twenty-four hours in the presence or absence of 10 μ M peptide at a density of 2.5 X 15 10⁵/well or 5 X 10⁵/well. After extensive washing (PBS 1X/0.005% Tween), biotinylated rat anti-mouse IFN γ antibody (PharMingen, Cat. No. 18112D, PharMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA) was added and incubated overnight at 4° C. For development, streptavidin-AKP (PharMingen, Cat. No. 13043E, PharMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA) and 1-Step™ NBT-BCIP development solution (Pierce, Cat. No. 34042, Pierce, P.O. Box 117, Rockford, IL 61105 USA) were added.
20

Pools of 20mer overlapping peptides encompassing the entire sequence of the HCV BK strain NS3 to NS5B were used to reveal HCV-specific IFN γ -secreting T cells. Similarly a single 20mer peptide encompassing a CD8+ epitope for
25 C57Black6 mice was used to detect CD8 response. Representative data from groups of C57Black6 and Balb/C mice (N=9-10) immunized with two injections of 25 or 50 μ g of plasmid vectors pV1Jns-NS, pV1Jns-NSmut and pV1Jns-NSOPTmut are shown in Figures 13A and 13B.

30 Example 4: Immunization of Rhesus Macaques

Rhesus macaques (N=3) were immunized by intramuscular injection with 5mg of plasmid pV1Jns-NSOPTmut in 7.5mg/ml CRL1005, Benzalkonium chloride 0.6 mM. Each animal received two doses in the deltoid muscle at 0, and 4 weeks.

CMI was measured at different time points by IFN- γ ELISPOT. This assay measures HCV antigen-specific CD8+ and CD4+ T lymphocyte responses, and can be used for a variety of mammals, such as humans, rhesus monkeys, mice, and rats.

5 The use of a specific peptide or a pool of peptides can simplify antigen presentation in CTL cytotoxicity assays, interferon-gamma ELISPOT assays and interferon-gamma intracellular staining assays. Peptides based on the amino acid sequence of various HCV proteins (core, E2, NS3, NS4A, NS4B, NS5A, NS5B) were prepared for use in these assays to measure immune responses in HCV DNA and
10 adenovirus vector vaccinated rhesus monkeys, as well as in HCV-infected humans. The individual peptides are overlapping 20-mers, offset by 10 amino acids. Large pools of peptides can be used to detect an overall response to HCV proteins while smaller pools and individual peptides may be used to define the epitope specificity of a response.

15

IFN γ ELISPOT

The IFN γ -ELISPOT assay provides a quantitative determination of HCV-specific T lymphocyte responses. PBMC are serially diluted and placed in microplate wells coated with anti-rhesus IFN- γ antibody (MD-1 U-Cytech). They are
20 cultured with a HCV peptide pool for 20 hours, resulting in the restimulation of the precursor cells and secretion of IFN- γ . The cells are washed away, leaving the secreted IFN bound to the antibody-coated wells in concentrated areas where the cells were sitting. The captured IFN is detected with biotinylated anti-rhesus IFN antibody (detector Ab U-Cytech) followed by alkaline phosphatase-conjugated streptavidin
25 (Pharmingen 13043E). The addition of insoluble alkaline phosphatase substrate results in dark spots in the wells at the sites where the cells were located, leaving one spot for each T cell that secreted IFN- γ .

The number of spots per well is directly related to the precursor frequency of antigen-specific T cells. Gamma interferon was selected as the cytokine
30 visualized in this assay (using species specific anti-gamma interferon monoclonal antibodies) because it is the most common, and one of the most abundant cytokines synthesized and secreted by activated T lymphocytes. For this assay, the number of spot forming cells (SFC) per million PBMCs is determined for samples in the

presence and absence (media control) of peptide antigens. Data from Rhesus macaques on PBMC from post dose two material are shown in Table 4.

Table 4

Pep pools	PV1J-NSOPTmut		
	21G	99C161	99C166
F (NS3p)	8	10	170
G (NS3h)	7	592	229
H (NS4)	3	14	16
I (NS5a)	5	71	36
L (NS5b)	14	23	11
M (NS5b)	3	35	8
DMSO	2	4	5

5 INF γ ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 5 mg DNA/dose in OPTIVAX/BAK of plasmid pV1Jns-NSOPTmut. Data are expressed as SFC7 10⁶ PBMC.

Example 5: Construction of Ad6 Pre-Adenovirus Plasmids

Ad6 pre-adenovirus plasmids were obtained as follows:

10

Construction of pAd6 E1-E3+ Pre-adenovirus Plasmid

An Ad6 based pre-adenovirus plasmid which can be used to generate first generation Ad6 vectors was constructed either taking advantage of the extensive sequence identity (approx. 98%) between Ad5 and Ad6 or containing only Ad6 regions. Homologous recombination was used to clone wtAd6 sequences into a bacterial plasmid.

15 A general strategy used to recover pAd6E1-E3+ as a bacterial plasmid containing Ad5 and Ad6 regions is illustrated in Figure 10. Cotransformation of BJ 5183 bacteria with purified wt Ad6 viral DNA and a second DNA fragment termed 20 the Ad5 ITR cassette resulted in the circularization of the viral genome by homologous recombination. The ITR cassette contains sequences from the right (bp 33798 to 35935) and left (bp 1 to 341 and bp 3525 to 5767) end of the Ad5 genome separated by plasmid sequences containing a bacterial origin of replication and an ampicillin resistance gene. The ITR cassette contains a deletion of E1 sequences from

Ad5 342 to 3524. The Ad5 sequences in the ITR cassette provide regions of homology with the purified Ad6 viral DNA in which recombination can occur.

- Potential clones were screened by restriction analysis and one clone was selected as pAd6E1-E3+. This clone was then sequenced in its entirety. pAd6E1-E3+ contains Ad5 sequences from bp 1 to 341 and from bp 3525 to 5548, Ad6 bp 5542 to 33784, and Ad5 bp 33967 to 35935 (bp numbers refer to the wt sequence for both Ad5 and Ad6). pAd6E1-E3+ contains the coding sequences for all Ad6 virion structural proteins which constitute its serotype specificity.

A general strategy used to recover pAd6E1-E3+ as a bacterial plasmid containing Ad6 regions is illustrated in Figure 11. Cotransformation of BJ 5183 bacteria with purified wt Ad6 viral DNA and a second DNA fragment termed the Ad6 ITR cassette resulted in the circularization of the viral genome by homologous recombination. The ITR cassette contains sequences from the right (bp 35460 to 35759) and left (bp 1 to 450 and bp 3508 to 3807) end of the Ad6 genome separated by plasmid sequences containing a bacterial origin of replication and an ampicillin resistance gene. These three segments were generated by PCR and cloned sequentially into pNEB193, generating pNEBAd6-3 (the ITR cassette). The ITR cassette contains a deletion of E1 sequences from Ad5 451 to 3507. The Ad6 sequences in the ITR cassette provide regions of homology with the purified Ad6 viral DNA in which recombination can occur.

Construction of pAd6 E1-E3- pre-adenovirus plasmids

Ad6 based vectors containing A5 regions and deleted in the E3 region were constructed starting with pAd6E1-E3+ containing Ad5 regions. A 5322 bp subfragment of pAd6E1-E3+ containing the E3 region (Ad6 bp 25871 to 31192) was subcloned into pABS.3 generating pABSAd6E3. Three E3 deletions were then made in this plasmid generating three new plasmids pABSAd6E3(1.8Kb) (deleted for Ad6 bp 28602 to 30440), pABSAd6E3(2.3Kb) (deleted for Ad6 bp 28157 to 30437) and pABSAd6E3(2.6Kb) (deleted for Ad6 bp 28157 to 30788). Bacterial recombination was then used to substitute the three E3 deletions back into pAd6E1-E3+ generating the Ad6 genome plasmids pAd6E1-E3-1.8Kb, pAd6E1-E3-2.3Kb and pAd6E1-E3-2.6Kb.

Example 6: Generation of Ad5 Genome Plasmid with the NS Sequence

A pcDNA3 plasmid (Invitrogen) containing the coding region NS3-NS4A-NS4B-NS5A was digested with *XmnI* and *NruI* restriction sites and the DNA fragment containing the CMV promoter, the NS3-NS4A-NS4B-NS5A coding sequence and the Bovine Growth Hormone (BGH) polyadenylation signal was cloned into the unique *EcoRV* restriction site of the shuttle vector pDelE1Spa, generating the Sva3-5A vector.

5 A pcDNA3 plasmid containing the coding region NS3-NS4A-NS4B-NS5A-NS5B was digested with *XmnI* and *EcoRI* (partial digestion), and the DNA fragment containing part of NS5A, NS5B gene and the BGH polyadenylation signal was cloned into the Sva3-5A vector, digested *EcoRI* and *BglII* blunted with Klenow, 10 generating the Sva3-5B vector.

The Sva3-5B vector was finally digested *SspI* and *Bst*1107I restriction sites and the DNA fragment containing the expression cassette (CMV promoter, NS3-NS4A-NS4B-NS5A-NS5B coding sequence and the BGH polyadenylation signal) flanked by adenovirus sequences was co-transformed with pAd5HVO (E1-,E3-) *ClaI* 15 linearized genome plasmid into the bacterial strain BJ5183, to generate pAd5HVONS. pAd5HVO contains Ad5 bp 1 to 341, bp 3525 to 28133 and bp 30818 to 35935.

Example 7: Generation of Adenovirus Genome Plasmids with the NSmut Sequence

Adenovirus genome plasmids containing an NS-mut sequence were generated in an Ad5 or Ad6 background. The Ad6 background contained Ad5 regions 20 at bases 1 to 450, 3511 to 5548 and 33967 to 35935.

pV1JNS3-5Akozak was digested with *BglII* and *XbaI* restriction enzymes and the DNA fragment containing the Kozak sequence and the sequence coding NS3-NS4A-NS4B-NS5A was cloned into a *BglII* and *XbaI* digested polypMRKpdelE1 shuttle vector. The resulting vector was designated shNS3- 25 5Akozak.

PolypMRKpdelE1 is a derivative of RKpdelE1(Pac/pIX/pack450) + CMVmin+BGHpA(str.) modified by the insertion of a polylinker containing recognition sites for *BglII*, *PmeI*, *SwaI*, *XbaI*, *SalI*, into the unique *BglII* restriction site present downstream the CMV promoter. MRKpdelE1(Pac/pIX/pack450) + 30 CMVmin + BGHpA(str.) contains Ad5 sequences from bp 1 to 5792 with a deletion of E1 sequences from bp 451 to 3510. The human CMV promoter and BGH polyadenylation signal were inserted into the E1 deletion in an E1 parallel orientation with a unique *BglII* site separating them.

The NS5B fragment, mutated to abrogate enzymatic activity and with a strong translation termination at the 3' end, was obtained by assembly PCR and inserted into the shNS3-5Akozak vector via homologous recombination, generating polypMRKpdelE1NSmut. In polypMRKpdelE1NSmut the NS-mut coding sequence 5 is under the control of CMV promoter and the BGH polyadenylation signal is present downstream.

The gene expression cassette and the flanking regions which contain adenovirus sequences allowing homologous recombination were excised by digestion with *PacI* and *Bst*1107I restriction enzymes and co-transformed with either 10 pAd5HVO (E1-,E3-) or pAd6E1-E3-2.6Kb *Cla*I linearized genome plasmids into the bacterial strain BJ5183, to generate pAd5HVONSmut and pAd6E1-,E3-NSmut, respectively.

pAd6E1-E3-2.6Kb contains Ad5 bp 1 to 341 and from bp 3525 to 5548, Ad6 bp 5542 to 28157 and from bp 30788 to 33784, and Ad5 bp 33967 to 15 35935 (bp numbers refer to the wt sequence for both Ad5 and Ad6). In both plasmids the viral ITR's are joined by plasmid sequences that contain the bacterial origin of replication and an ampicillin resistance gene.

Example 8: Generation of Adenovirus Genome Plasmids with the NSOPTmut

The human codon-optimized synthetic gene (NSOPTmut) provided by 20 SEQ. ID. NO. 3 cloned into a pCRBlunt vector (Invitrogen) was digested with *Bam*H1 and *Sal*I restriction enzymes and cloned into *Bgl*II and *Sal*I restriction sites present in the shuttle vector polypMRKpdelE1. The resulting clone (polypMRKpdelE1NSOPTmut) was digested with *Pac*I and *Bst*1107I restriction enzymes and co-transformed with either pAd5HVO (E1-,E3-) or pAd6E1-E3-2.6Kb 25 *Cla*I linearized genome plasmids, into the bacterial strain BJ5183, to generate pAd5HVONSOPTmut and pAd6E1-,E3-NSOPTmut, respectively.

Example 9: Rescue and Amplification of Adenovirus Vectors

Adenovectors were rescued in Per.6 cells. Per.C6 were grown in 10% 30 FCS / DMEM supplemented by L-glutamine (final 4mM), penicillin/streptomycin (final 100 IU/ml) and 10 mM MgCl₂. After infection, cells were kept in the same medium supplemented by 5% horse serum (HS). For viral rescue, 2.5 X 10⁶ Per.C6 were plated in 6 cm ø Petri dishes.

Twenty-four hours after plating, cells were transfected by calcium phosphate method with 10 µg of the *Pac I* linearized adenoviral DNA. The DNA precipitate was left on the cells for 4 hours. The medium was removed and 5% HS/DMEM was added.

5 Cells were kept in a CO₂ incubator until a cytopathic effect was visible (1 week). Cells and supernatant were recovered and subjected to 3X freeze/thawing cycles (liquid nitrogen / water bath at 37°C). The lysate was centrifuged at 3000 rpm at - 4°C for 20 minutes and the recovered supernatant (corresponding to a cell lysate containing virus passed on cells only once; P1) was used, in the amount of 1 ml/ dish,
10 to infect 80-90% confluent Per.C6 in 10 cm ø Petri dishes. The infected cells were incubated until a cytopathic effect was visible, cells and supernatant recovered and the lysate prepared as described above (P2).

15 P2 lysate (4 ml) were used to infect 2 X 15 cm ø Petri dishes. The lysate recovered from this infection (P3) was kept in aliquots at -80°C as a stock of virus to be used as starting point for big viral preparations. In this case, 1 ml of the stock was enough to infect 2 X 15 cm ø Petri dishes and resulting lysate (P4) was used for the infection of the Petri dishes devoted to the large scale infection.

20 Further amplification was obtained from the P4 lysate which was diluted in medium without FCS and used to infect 30 X 15 cm ø Petri dishes (with Per.C6 80%-90% confluent) in the amount of 10 ml/dish. Cells were incubated 1 hour in the CO₂ incubator, mixing gently every 20 minutes. 12 ml / dish of 5% HS / DMEM was added and cells were incubated until a cytopathic effect was visible (about 48 hours).

25 Cells and supernatant were collected and centrifuged at 2K rpm for 20 minutes at 4°C. The pellet was resuspended in 15 ml of 0.1 M Tris pH=8.0. Cells were lysed by 3X freeze/thawing cycles (liquid nitrogen / water bath at 37°C). 150 µl of 2 M MgCl₂ and 75 µl of DNase (10 mg of bovine pancreatic deoxyribonuclease I in 10 ml of 20 mM Tris-HCl pH= 7.4, 50 mM NaCl, 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, 50% glycerol) were added. After a 1 hour incubation at 37°C
30 in a water bath (vortex every 15 minutes) the lysate was centrifuged at 4K rpm for 15 minutes at 4°C. The recovered supernatant was ready to be applied on CsCl gradient.

The CsCl gradients were prepared in SW40 ultra-clear tubes as follows:

- 0.5 ml of 1.5d CsCl
35 3 ml of 1.35d CsCl

3 ml of 1.25d CsCl

5-ml/ tube of viral supernatant was applied.

If necessary, the tubes were topped up with 0.1 M tris-Cl pH=8.0.

Tubes were centrifuged at 35K rpm for 1 hour at -10⁰C with rotor SW40. The viral
5 bands (located at the 1.25/1.35 interface) were collected using a syringe.

The virus was transferred into a new SW40 ultraclear tube and 1.35d
CsCl was added to top the tube up. After centrifugation at 35K rpm for 24 hours at
10⁰C in the rotor SW40, the virus was collected in the smallest possible volume and
dialyzed extensively against buffer A105 (5 mM Tris, 5% sucrose, 75 mM NaCl, 1
10 mM MgCl₂, 0.005% polysorbate 80 pH=8.0). After dialysis, glycerol was added to
final 10% and the virus was stored in aliquots at - 80⁰C.

Example 10: Enhanced Adenovector Rescue

First generation Ad5 and Ad6 vectors carrying HCV NSOPTmut
15 transgene were found to be difficult to rescue. A possible block in the rescue process
might be attributed to an inefficient replication of plasmid DNA that is a sub-optimal
template for the replication machinery of adenovirus. The absence of the terminal
protein linked to the 5'ends of the DNA (normally present in the viral DNA),
associated with the very high G-C content of the transgene inserted in the E1 region of
20 the vector, may be causing a substantial reduction in replication rate of the plasmid-
derived adenovirus.

To set up a more efficient and reproducible procedure for rescuing Ad
vectors, an expression vector (pE2; Figure 19) containing all E2 proteins (polymerase,
pre-terminal protein and DNA binding protein) as well as E4 orf6 under the control of
25 tet-inducible promoter was employed. The transfection of pE2 in combination with a
normal preadeno plasmid in PerC6 and in 293 leads to a strong increase of Ad DNA
replication and to a more efficient production of complete infectious adenovirus
particles.

30 *Plasmid Construction*

pE2 is based on the cloning vector pBI (CLONTECH) with the
addition of two elements to allow episomal replication and selection in cell culture:
(1) the EBV-OriP (EBV [nt] 7421-8042) region permitting plasmid replication in
synchrony with the cell cycle when EBNA-1 is expressed and (2) the hygromycin-B
35 phosphotransferase (HPH)-resistance gene allowing a positive selection of

transformed cells. The two transcriptional units for the adenoviral genes E2 a and b and E4-Orf6 were constructed and assembled in pE2 as described below.

The Ad5-Polymerase *Clal/SphI* fragment and the Ad5-pTP *Acc65/EcoRV* fragment were obtained from pVac-Pol and pVac-pTP (Stunnenberg *et al.* NAR 16:2431-2444, 1988). Both fragments were filled with Klenow and cloned into the *Sall* (filled) and *EcoRV* sites of pBI, respectively obtaining pBI-Pol/pTP.

EBV-OriP element from pCEP4 (Invitrogen) was first inserted within two chicken β-globin insulator dimers by cloning it into *BamHI* site of pJC13-1 (Chung *et al.*, Cell 74(3):505-14, 1993). HS4-OriP fragment from pJC13-OriP was 10 then cloned inside pSA1mv (a plasmid containing tk-Hygro-B resistance gene expression cassette as well as Ad5 replication origin), the ITR's arranged as head-to-tail junction, obtained by PCR from pFG140 (Graham, EMBO J. 3:2917-2922, 1984) using the following primers: 5'-TCGAATCGATA CGCGAACCTACGC-3' (SEQ. ID. NO. 16) and 5'-TCGACGTGTCGACTTCGAAGCGCACACCAAAACGTC-3' (SEQ. ID. NO. 17), thus generating pMVHS4Orip. A DNA fragment from pMVHS4Orip, containing the insulated OriP, Ad5 ITR junction and tk-HygroB cassette, was then inserted into pBI-Pol/pTP vector restricted *AseI/AatII* generating pBI-Pol/pTPHS4 .

To construct the second transcriptional unit expressing Ad5-Orf6 as 20 well as Ad5-DBP, E4orf6 (Ad 5 [nt] 33193-34077) obtained by PCR was first inserted into pBI vector, generating pBI-Orf6. Subsequently, DBP coding DNA sequence (Ad 5 [nt] 22443-24032) was inserted into pBI-Orf6 obtaining the second bi-directional Tet-regulated expression vector (pBI-DBP/E4orf6). The original polyA signals present in pBI were substituted with BGH and SV40 polyA.

pBI-DBP/E4orf6 was then modified by inserting a DNA fragment 25 containing the Adeno5-ITRs arranged in head-to-tail junction plus the hygromycin B resistance gene obtained from plasmid pSA-1mv. The new plasmid pBI-DBP/E4orf6shuttle was then used as donor plasmid to insert the second tet-regulated transcriptional unit into pBI-Pol/pTPHS4 by homologous recombination using *E. coli* strain BJ5183 obtaining pE2.

Cell lines, Transfections and Virus Amplification

PerC6 cells were cultured in Dulbecco's modified Eagle's Medium (DMEM) plus 10% fetal bovine serum (FBS), 10 mM MgCl₂, penicillin (100 U/ml), 35 streptomycin (100 µg/ml) and 2 mM glutamine.

All transient transfections were performed using Lipofectamine2000 (Invitrogen) as described by the manufacturer. 90% confluent PERC.6™ planted in 6-cm plates were transfected with 3.5 µg of Ad5/6NSOPTmut pre-adeno plasmids, digested with PacI, alone or in combination with 5 µg pE2 plus 1 µg pUHD52.1.

5 pUHD52.1 is the expression vector for the reverse tet transactivator 2 (rtTA2) (Urlinger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 97(14):7963-7968, 2000). Upon transfection, cells were cultivated in the presence of 1 µg/ml of doxycycline to activate pE2 expression. 7 days post-transfection cells were harvested and cell lysate was obtained by three cycles of freeze-thaw. Two ml of cell lysate were used to infect

10 a second 6-cm dish of PerC6. Infected cells were cultivated until a full CPE was observed then harvested. The virus was serially passaged five times as described above, then purified on CsCl gradient. The DNA structure of the purified virus was controlled by endonuclease digestion and agarose gel electrophoresis analysis and compared to the original pre-adeno plasmid restriction pattern.

15

Example 11: Partial Optimization of HCV Polyprotein Encoding Nucleic acid

Partial optimization of HCV polyprotein encoding nucleic acid was performed to facilitate the production of adenovectors containing codons optimized for expression in a human host. The overall objective was to provide for increased

20 expression due to codon optimization, while facilitating the production of an adenovector encoding HCV polyprotein.

Several difficulties were encountered in producing an adenovector encoding HCV polyprotein with codons optimized for expression in a human host. An adenovector containing an optimized sequence (SEQ. ID. NO. 3) was found to be

25 more difficult to synthesize and rescue than an adenovector containing a non-optimized sequence (SEQ. ID. NO. 2).

The difficulties in producing an adenovector containing SEQ. ID. NO. 3 were attributed to a high GC content. A particularly problematic region was the region at about position 3900 of NSOPTmut (SEQ. ID. NO. 3).

30 Alternative versions of optimized HCV encoding nucleic acid sequence were designed to facilitate its use in an adenovector. The alternative versions, compared to NSOPTmut, were designed to have a lower overall GC content, to reduce/avoid the presence of potentially problematic motifs of consecutive G's or C's, while maintaining a high level of codon optimization to allow improved

35 expression of the encoded polyprotein and the individual cleavage products.

A starting point for the generation of a suboptimally codon-optimized sequence is the coding region of the NSOPTmut nucleotide sequence (bases 7 to 5961 of SEQ. ID. NO. 3). Values for codon usage frequencies (normalized to a total of 1.0 for each amino acid) were taken from the file human_high.cod available in the
5 Wisconsin Package Version 10.3 (Accelrys Inc., a wholly owned subsidiary of Pharmacopeia, Inc.).

To reduce the local and overall GC content a table defining preferred codon substitutions for each amino acid was manually generated. For each amino acid the codon having 1) a lower GC content as compared to the most frequent codon and
10 2) a relatively high observed codon usage frequency (as defined in human_high.cod) was chosen as the replacement codon. For example for Arg the codon with the highest frequency is CGC. Out of the other five alternative codons encoding Arg (CGG, AGG, AGA, CGT, CGA) three (AGG, CGT, CGA) reduce the GC content by 1 base, one (AGA) by two bases and one (CGG) by 0 bases. Since the AGA codon is
15 listed in human_high.cod as having a relatively low usage frequency (0.1), the codon substituting CGC was therefore chosen to be AGG with a relative frequency of 0.18. Similar criteria were applied in order to establish codon replacements for the other amino acids resulting in the list shown in Table 5. Parameters applied in the following optimization procedure were determined empirically such that the resulting sequence
20 maintained a considerably improved codon usage (for each amino acid) and the GC content (overall and in form of local stretches of consecutive G's and/or C's) was decreased.

Two examples of partial optimized HCV encoding sequences are provided by SEQ. ID. NO. 10 and SEQ. ID. NO. 11. SEQ. ID. NO. 10 provides a
25 HCV encoding sequence that is partially optimized throughout. SEQ. ID. NO. 11 provides an HCV encoding sequence fully optimized for codon usage with the exception of a region that was partially optimized.

Codon optimization was performed using the following procedure:
Step 1) The coding region of the input fully optimized NSOPTmut
30 sequence was analyzed using a sliding window of 3 codons (9 bases) shifting the window by one codon after each cycle. Whenever a stretch containing 5 or more consecutive C's and/or G's was detected in the window the following replacement rule was applied: Let N indicate the number of codon replacements previously performed. If N is odd replace the middle codon in the window with the codon specified in Table
35 5, if N is even replace the third terminal codon in the window with the codon

specified in a codon optimization table such as human_high.cod. If Leu or Val is present at the second or third codon do not apply any replacement in order not to introduce Leu or Val codons with very low relative codon usage frequency (see, for example, human_high.cod). In the following cycle analysis of the shifted window was 5 then applied to a sequence containing the replacements of the previous cycle.

The alternating replacement of the middle and terminal codon in the 3 codon window was found empirically to give a more satisfying overall maintenance of optimized codon usage while also reducing GC content (as judged from the final sequence after the procedure). In general, however, the precise replacement strategy 10 depends on the amino acid sequence encoded by the nucleotide sequence under analysis and will have to be determined empirically.

Step 2) The sequence containing all the codon replacements performed during step 1) was then subjected to an additional analysis using a sliding window of 21 codons (63 bases) in length: according to an adjustable parameter the overall GC 15 content in the window was determined. If the GC content in the window was higher than 70% the following codon replacement strategy was applied: In the window replace the codons for the amino acids Asn, Asp, Cys, Glu, His, Ile, Lys, Phe, Tyr by the codons given in Table 5. Restriction of the replacement to this set of amino acids was motivated by the fact that a) the replacement codon still has an acceptably high 20 frequency of usage in human_high.cod and b) the average overall human codon usage in CUTG for the replacement codon is nearly as high as the most frequent codon. In the following cycle analysis of the shifted window is then applied to a sequence containing the replacements of the previous cycle.

The threshold 70% was determined empirically by compromising 25 between an overall reduction in GC content and maintenance of a high codon optimization for the individual amino acids. As in step 1) the precise replacement strategy (choice of amino acids and GC content threshold value) will again depend on the amino acid sequence encoded by the nucleotide sequence under analysis and will have to be determined empirically.

Step 3) The sequence generated by steps 1) and 2) was then manually 30 edited and additional codons were changed according to the following criteria: Regions still having a GC content higher than 70% over a window of 21 codons were examined manually and a few codons were replaced again following the scheme given in Table 5.

Subsequent steps were performed to provide for useful restriction sites, remove possible open reading frames on the complementary strand, to add homologous recombinant regions, to add a Kozac signal, and to add a terminator. These steps are numbered 4-7

5 Step 4) The sequence generated in step 3 was examined for the absence of certain restriction sites (BglII, PmeI and XbaI) and presence of only 1 StuI site to allow a subsequent cloning strategy using a subset of restriction enzymes. Two sites (one for BglII and one for StuI) were removed from the sequence by replacing codons that were part of the respective recognition sites.

10 Step 5) The sequence generated by steps 1) through 4) was then modified according to allow subsequent generation of a modified NSOPTmut sequence (by homologous recombination). In the sequence obtained from steps 1) through 4) the segment comprising base 3556 to 3755 and the segment comprising base 4456 to 4656 were replaced by the corresponding segments from NSOPTmut.

15 The segment comprising bases 3556 to 4656 of SEQ. ID. NO. 10 can be used to replace the problematic region in NSOPTmut (around position 3900) by homologous recombination thus creating the variant of NSOPTmut having the sequence of SEQ. ID. NO. 11.

Step 6) Analysis of the sequence generated through steps 1) to 5) revealed a potential open reading frame spanning nearly the complete fragment on the complementary strand. Removal of all codons CTA and TTA (Leu) and TCA (Ser) from the sense strand effectively removed all stop codons in one of the reading frames on the complementary strand. Although the likelihood for transcription of this complementary strand open reading frame and subsequent translation into protein is very small, in order to exclude a potential interference with the transcription and subsequent translation of the sequence encoded on the sense strand, TCA codons for Ser were introduced on the sense approximately every 500 bases. No changes were introduced in the segments introduced during step 5) to allow homologous recombination. The TCA codon for Ser was preferred over the CTA and TTA codons for Leu because of the higher relative frequency for TCA (0.05) as compared to CTA (0.02) and TTA (0.03) in human_high.cod. In addition, the average human codon usage from CUTG favored TCA (0.14 against 0.07 for CTA and TTA).

Step 7) In a final step GCCACC was added at the 5' end of the sequence to generate an optimized internal ribosome entry site (Kozak signal) and a 35 TAAA stop signal was added at the 3'. To maintain the initiation of translation

properties of NSsuboptmut the first 8 codons of the coding region were kept identical to the NSOPTmut sequence. The resulting sequence was again checked for the absence of BglII, PmeI and XbaI recognition sites and the presence of only 1 StuI site.

- The NSsuboptmut sequence (SEQ. ID. NO. 10) has an overall reduced
 5 GC content (63.5%) as compared to NSOPTmut (70.3%) and maintains a well optimized level of codon usage optimization. Nucleotide sequence identity of NSsuboptmut is 77.2% with respect to NSmut.

Table 5: Definition of codon replacements performed during steps 1) and 2).

10

Amino Acid	Most frequent codon	Relative frequency	Reduction in GC content (bases)	Replacement codon	Relative frequency
Amino Acids where the replacement codon reduces the codon GC-content by 1 base					
Ala	GCC	0.51	1	GCT	0.17
Arg	CGC	0.37	1	AGG	0.18
Asn	AAC	0.78	1	AAT	0.22
Asp	GAC	0.75	1	GAT	0.25
Cys	TGC	0.68	1	TGT	0.32
Glu	GAG	0.75	1	GAA	0.25
Gln	CAG	0.88	1	CAA	0.12
Gly	GGC	0.50	1	GGA	0.14
His	CAC	0.79	1	CAT	0.21
Ile	ATC	0.77	1	ATT	0.18
Lys	AAG	0.82	1	AAA	0.18
Phc	TTC	0.80	1	TTT	0.20
Pro	CCC	0.48	1	CCT	0.19
Ser	AGC	0.34	1	TCT	0.13
Thr	ACC	0.51	1	ACA	0.14
Tyr	TAC	0.74	1	TAT	0.26
Amino Acids with no alternative codon					
Met	ATG	1.00	0	ATG	1.00
Trp	TGG	1.00	0	TGG	1.00

Amino Acids where the replacement codon has a very low relative frequency. These amino acids were excluded from the replacement procedure					
Leu	CTG	0.58	1	TTG	0.06
Val	GTG	0.64	1	GTT	0.07

Example 12: Virus Characterization

Adenovectors were characterized by: (a) measuring the physical particles/ml; (b) running a TaqMan PCR assay; and (c) checking protein expression
5 after infection of HeLa cells.

a) Physical Particles Determination

CsCl purified virus was diluted 1/10 and 1/100 in 0.1% SDS PBS. As a control, buffer A105 was used. These dilutions were incubated 10 minutes at 55°C.
10 After spinning the tubes briefly, O.D. at 260 nm was measured. The amount of viral particles was calculated as follows: 1 OD 260 nm = 1.1 X 10¹² physical particles/ml. The results were typically between 5 X 10¹¹ and 1 X 10¹² physical particles /ml.

b) TaqMan PCR Assay

15 TaqMan PCR assay was used for adenovectors genome quantification (Q-PCR particles/ml). TaqMan PCR assay was performed using the ABI Prism 7700-sequence detector. The reaction was performed in a final 50 µl volume in the presence of oligonucleotides (at final 200 nM) and probe (at final 200 µM) specific for the adenoviral backbone. The virus was diluted 1/10 in 0.1% SDS PBS and
20 incubated 10 minutes at 55°C. After spinning the tube briefly, serial 1/10 dilutions (in water) were prepared. 10 µl the 10⁻³, 10⁻⁵ and 10⁻⁷ dilutions were used as templates in the PCR assay.

The amount of particles present in each sample was calculated on the basis of a standard curve run in the same experiment. Typically results were between
25 1 X 10¹² and 3 X 10¹² Q-PCR particles /ml.

c) Expression of HCV Non-Structural Proteins

Expression of HCV NS proteins was tested by infection of HeLa cells. Cells were plated the day before the infection at 1.5 X 10⁶ cells/dish (10 cm ø Petri
30 dishes). Different amounts of CsCl purified virus corresponding to m.o.i. of 50, 250

and 1250 pp/cell were diluted in medium (FCS free) up to a final volume of 5 ml. The diluted virus was added on the cells and incubated for 1 hour at 37°C in a CO₂ incubator (gently mixing every 20 minutes). 5 ml of 5% HS-DMEM was added and the cells were incubated at 37°C for 48 hours.

- 5 Cell extracts were prepared in 1% Triton/TEN buffer. The extracts were run on 10% SDS-acrylamide gel, blotted on nitrocellulose and assayed with antibodies directed against NS3, NS5a and NS5b in order to check the correct polyprotein cleavage. Mock-infected cells were used as a negative control. Results from representative experiments testing the Ad5-NS, MRKAd5-NSmut, MRKAd6-10 NSmut and MRKAd6-NSOPTmut are shown in Figure 14.

Example 13: Mice Immunization with Adenovectors Encoding Different NS Cassettes

- 15 The adenovectors Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut and MRKAd6-NSOPTmut were injected in C57Black6 mice strains to evaluate their potential to elicit anti-HCV immune responses. Groups of animals (N=9-10) were injected intramuscularly with 10⁹ pp of CsCl purified virus. Each animal received two doses at three weeks interval.

- 20 Humoral immune response against the NS3 protein was measured in post dose two sera from C57Black6 immunized mice by ELISA on bacterially expressed NS3 protease domain. Antibodies specific for the tested antigen were detected with geometric mean titers (GMT) ranging from 100 to 46000 (Tables 6, 7, 8 and 9).

25

Table 6: Ad5-NS

Mice n.	1	2	3	4	5	6	7	8	9	10	GMT
Titer	50	253	50	50	50	2257	504	50	50	50	108

30

Table 7: Ad5-NSmut

											GMT
Mice n.	11	12	13	14	15	16	17	18	19	20	
Titer	3162	78850	87241	6796	12134	3340	18473	13093	76167	49593	23645

Table 8: MRKAd6-NSmut

5

											GMT
Mice n.	21	22	23	24	25	26	27	28	29	30	
Titer	125626	39751	40187	65834	60619	69933	21555	49348	29290	26859	46461

Table 9: MRKAd6-NSOPTmut

								GMT
Mice n.	31	32	33	34	35	36	37	
Titer	25430	3657	893	175	10442	49540	173	2785

10 T cell response in C57Black6 mice was analyzed by the quantitative ELISPOT assay measuring the number of IFN γ secreting T cells in response to five pools (named from F to L+M) of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence. Specific CD8+ response induced in C57Black6 mice was analyzed by the same assay using a 20mer peptide 15 encompassing a CD8+ epitope for C57Black6 mice (pep1480). Cells secreting IFN γ in an antigen specific-manner were detected using a standard ELispot assay.

15 Spleen cells, splenocytes and peptides were produced and treated as described in Example 3, *supra*. Representative data from groups of C57Black6 mice (N=9-10) immunized with two injections of 10^9 viral particles of vectors Ad5-NS, 20 MRKAd5-NSmut and MRKAd6-NSmut are shown in Figure 15.

Example 14: Immunization of Rhesus macaques with Adenovectors

Rhesus macaques (N=3-4) were immunized by intramuscular injection of CsCl purified Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut or MRKAd6-

NSOPTmut virus. Each animal received two doses of 10^{11} or 10^{10} vp in the deltoid muscle at 0, and 4 weeks.

CMI was measured at different time points by a) IFN- γ ELISPOT (see Example 3, *supra*), b) IFN- γ ICS and c) bulk CTL assays. These assays measure HCV antigen-specific CD8+ and CD4+ T lymphocyte responses, and can be used for a variety of mammals, such as humans, rhesus monkeys, mice, and rats.

The use of a specific peptide or a pool of peptides can simplify antigen presentation in CTL cytotoxicity assays, interferon-gamma ELISPOT assays and interferon-gamma intracellular staining assays. Peptides based on the amino acid sequence of various HCV proteins (core, E2, NS3, NS4A, NS4B, NS5a, NS5b) were prepared for use in these assays to measure immune responses in HCV DNA and adenovirus vector vaccinated rhesus monkeys, as well as in HCV-infected humans. The individual peptides are overlapping 20-mers, offset by 10 amino acids. Large pools of peptides can be used to detect an overall response to HCV proteins while smaller pools and individual peptides may be used to define the epitope specificity of a response.

IFN- γ ICS

For IFN- γ ICS, 2×10^6 PBMC in 1 ml R10 (RPMI medium, 20 supplemented with 10% FCS) were stimulated with peptide pool antigens. Final concentration of each peptide was 2 μ g/ml. Cells were incubated for 1 hour in a CO₂ incubator at 37°C and then Brefeldin A was added to a final concentration of 10 μ g /ml to inhibit the secretion of soluble cytokines. Cells were incubated for additional 14-16 hours at 37°C.

Stimulation was done in the presence of co-stimulatory antibodies: CD28 and CD49d (anti-humanCD28 BD340975 and anti-humanCD49d BD340976). After incubation, cells were stained with fluorochrome-conjugated antibodies for surface antigens: anti-CD3, anti-CD4, anti-CD8 (CD3-APC Biosource APS0301, CD4-PE BD345769, CD8-PerCP BD345774).

To detect intracellular cytokines, cells were treated with FACS permeabilization buffer 2 (BD340973), 2x final concentration. Once fixed and permeabilized, cells were incubated with an antibody against human IFN- γ , IFN- γ FITC (Biosource AHC4338).

Cells were resuspended in 1% formaldehyde in PBS and analyzed at 35 FACS within 24 hours. Four color FACS analysis was performed on a FACSCalibur

instrument (Becton Dickinson) equipped with two lasers. Acquisition was done gating on the lymphocyte population in the Forward versus Side Scatter plot coupled with the CD3, CD8 positive populations. At least 30,000 events of the gate were taken. The positive cells are expressed as number of IFN- γ expressing cells over 10^6

5 lymphocytes.

IFN- γ ELISPOT and IFN- γ ICS data from immunized monkeys after one or two injections of 10^{10} or 10^{11} vp of the different adenovectors are reported in Figures 16A-16D, 17A, and 17B.

10 *Bulk CTL Assays*

A distinguishing effector function of T lymphocytes is the ability of subsets of this cell population to directly lyse cells exhibiting appropriate MHC-associated antigenic peptides. This cytotoxic activity is most often associated with CD8+ T lymphocytes.

15 PBMC samples were infected with recombinant vaccine viruses expressing HCV antigens *in vitro* for approximately 14 days to provide antigen restimulation and expansion of memory T cells. Cytotoxicity against autologous B cell lines treated with peptide antigen pools was tested.

The lytic function of the culture is measured as a percentage of specific lysis resulted from chromium released from target cells during 4 hours incubation with CTL effector cells. Specific cytotoxicity is measured and compared to irrelevant antigen or excipient-treated B cell lines. This assay is semi-quantitative and is the preferred means for determining whether CTL responses were elicited by the vaccine. Data after two injections from monkeys immunized with 10^{11} vp/dose with adenovectors Ad5-NS, MRKAd5-NSmut and MRKAd6-NSmut are reported in Figures 18A-18F.

Other embodiments are within the following claims. While several embodiments have been shown and described, various modifications may be made without departing from the spirit and scope of the present invention.

WHAT IS CLAIMED IS:

1. A nucleic acid comprising a nucleotide sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ ID NO: 1, provided that said polypeptide has sufficient protease activity to process itself to produce an NS5B protein and said NS5B protein is enzymatically inactive.
2. The nucleic acid of claim 1, wherein said nucleotide sequence is substantially similar to the coding sequence of SEQ ID NO: 2.
10
3. The nucleic acid of claim 1, wherein said nucleotide sequence encodes for the polypeptide of SEQ ID NO: 1.
4. The nucleic acid of claim 3, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
15
5. The nucleic acid of claim 3, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2 or SEQ ID NO: 3.
20
6. The nucleic acid of any one of claims 1-5, wherein said nucleic acid is an expression vector capable of expressing said polypeptide from said nucleotide sequence in a human cell.
- 25
7. A nucleic acid comprising a gene expression cassette able to express a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ ID NO: 1 in a human cell, provided that said polypeptide can process itself to produce an NS5B protein and said NS5B protein is enzymatically inactive, said expression cassette comprising:
 - a) a promoter transcriptionally coupled to a nucleotide sequence encoding said polypeptide;
 - 30
 - b) a 5' ribosome binding site functionally coupled to said nucleotide sequence,

c) a terminator joined to the 3' end of said nucleotide sequence, and
d) a 3' polyadenylation signal functionally coupled to said nucleotide sequence.

5 8. The nucleic acid of claim 7, wherein said nucleotide sequence is substantially similar to either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

9. The nucleic acid of claim 8, wherein said nucleic acid is a
10 shuttle vector further comprising a selectable marker, an origin of replication, a first adenovirus homology region and a second adenovirus homology region flanking said expression cassette, wherein said first homology region has at least about 100 base pairs substantially homologous to at least right end of a wild-type adenovirus region from about base pairs 1-425, and said second homology region has at least about 100
15 base pairs substantially homologous to at least the left end of a wild-type adenovirus region from about base pairs 3511-5792 of Ad5 or corresponding region of another adenovirus.

10. The nucleic acid of claim 9, wherein said nucleotide sequence
20 encodes for a polypeptide of SEQ ID NO: 1.

11. The nucleic acid of claim 9, wherein said nucleotide sequence is SEQ ID NO: 2.

25 12. The nucleic acid of claim 9, wherein said nucleotide sequence is either SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

30 13. The nucleic acid of claim 8, wherein said nucleic acid is a plasmid suitable for administration into a human and further comprises a prokaryotic origin of replication and a gene coding for a selectable marker.

14. The nucleic acid of claim 13, wherein said nucleotide sequence encodes for a polypeptide of SEQ ID NO: 1.

15. The nucleic acid of claim 14, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

5 16. The nucleic acid of claim 14, wherein said nucleotide sequence is the coding sequence of SEQ ID NO: 2 or SEQ ID NO: 3.

10 17. The nucleic acid of claim 14, wherein said promoter is the human intermediate early cytomegalovirus promoter (intron A), said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the bovine growth hormone (BGH) polyadenylation signal.

15 18. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising a selectable marker, an origin of replication, and a recombinant adenovector genome containing an E1 deletion, an E3 deletion, and said expression cassette.

20 19. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising a selectable marker, an origin of replication, and

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) said gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to said first region;
- c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and

f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

5 20. The nucleic acid of claim 19, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

10 21. The nucleic acid of claim 20, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

15 22. The nucleic acid of claim 21, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

20 23. The nucleic acid of claim 19, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

25 24. The nucleic acid of claim 23, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

30 25. The nucleic acid of claim 24, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

35 26. The nucleic acid of claim 24, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2 or SEQ ID NO: 3.

27. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising an origin of replication, a selectable marker, and:

- 5 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- 10 c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- 15 e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base 20 pair 35759 corresponding to Ad6, joined to said fourth region.

28. The nucleic acid of claim 27, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region 25 corresponds to Ad5.

29. The nucleic acid of claim 28, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation 30 signal.

30. The nucleic acid of claim 27, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 of Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth 35 region corresponds to Ad5 or Ad6.

31. The nucleic acid of claim 30, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

32. The nucleic acid of claim 8, wherein said nucleic acid is a adenovector consisting of a nucleotide sequence substantially similar to of SEQ ID NO. 4 or a derivative thereof, wherein said derivative thereof has the HCV 10 polyprotein encoding sequence present in SEQ ID NO: 4 replaced with the HCV polyprotein encoding sequence of either SEQ ID NO: 3, SEQ ID NO: 10 or SEQ ID NO: 11.

33. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector having an adenovector genome containing an E1 deletion, an E3 deletion, and said expression cassette

34. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector consisting of:

20 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

 b) said gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to said first region;

 c) a second adenovirus region from about base pair 3511 to about 25 base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said expression cassette;

 d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;

30 e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and

 f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base 35 pair 35759 corresponding to Ad6, joined to said fourth region.

35. The nucleic acid of claim 34, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region
5 corresponds to Ad5.

36. The nucleic acid of claim 35, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation signal.
10

37. The nucleic acid of claim 36, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
15

38. The nucleic acid of claim 34, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.
20

39. The nucleic acid of claim 37, where said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation signal.
25

40. The nucleic acid of claim 39, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
30

41. The nucleic acid of claim 39, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is SEQ ID NO: 2 or SEQ ID NO: 3.
35

42. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector consisting of:
40

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

43. The nucleic acid of claim 42, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

44. The nucleic acid of claim 42, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

45. An adenovector consisting of the nucleic acid sequence of SEQ ID NO. 4 or a derivative thereof, wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ ID NO: 4 replaced with the HCV polyprotein encoding sequence of either SEQ ID NO: 3, SEQ ID NO: 10 or SEQ ID NO: 11.

46. An adenovector produced by a process comprising the steps of:

- a) producing an adenovirus genome plasmid by homologous recombination between the shuttle vector of claim 9 and a nucleic acid comprising:
 - a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- 5 a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 10 28156 corresponding to Ad6, joined to said second region;
- a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and
- a fifth adenovirus region from about base pair 33967 to about 15 base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region; and
- b) rescuing said adenovector from said adenovirus plasmid.

47. A cultured recombinant cell comprising the nucleic acid of 20 claim 6.

48. A cultured recombinant cell comprising the nucleic acid of any one of claims 9-46.

49. A method of making an adenovector comprising the steps of:
- a) producing an adenovirus genome plasmid comprising a gene expression cassette by homologous recombination between the nucleic acid of claim 9 and a nucleic acid comprising:
 - a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;

- a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- 5 a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and
- 10 a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region; and
- 10 b) rescuing said recombinant adenovirus from said recombinant adenovirus plasmid.

50. A pharmaceutical composition comprising the nucleic acid of any one of claims 13-17 and 32-46 and pharmaceutically acceptable carrier.
- 15
51. A method of treating a patient comprising the step of administering to said patient an effective amount of the nucleic acid of any one of claims 13-17 and 32-46.
- 20
52. The method of claim 51, wherein said patient is a human.
53. The method of claim 52, wherein said patient is not infected with HCV.
- 25
54. The method of claim 52, wherein said patient is infected with HCV.
55. A recombinant nucleic acid comprising one or more Ad6 regions and a region not present in Ad6, wherein at least one Ad6 region is selected from the group consisting of: E1A, E1B, E2B, E2A, E4, L1, L2, L4, and L5.
- 30
56. The recombinant nucleic acid of claim 55, wherein said region not present in Ad6, is an expression cassette coding for a polypeptide not found in Ad6.
- 35

57. The recombinant nucleic acid of claim 56, wherein said recombinant nucleic acid is an adenovirus vector defective in at least E1 that is able to replicate when E1 is supplied *in trans*.

5 58. The recombinant nucleic acid of claim 57, wherein said vector consists of:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) said gene expression cassette in an E1 parallel or E1 anti-parallel orientation joined to said first region;
- c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said gene expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- e) an optionally present fourth region from about base pair 28134 to about base pair 30817 corresponding to Ad5, or from about base pair 28157 to about 15 30789 corresponding to Ad6, joined to said third region;
- f) a fifth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, wherein said fifth region is joined to said fourth region if said fourth region is present, or said fifth is joined to said third region if said fourth region is not present; and
- g) a sixth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region;
provided that at least one of said second, third, and fifth regions is from Ad6.

20

30 59. The recombinant nucleic acid of claim 57, wherein said vector consists of:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
 - c) a third adenovirus region from about base pair 5549 to about 5 base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
 - d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
 - e) a fourth adenovirus region from about base pair 30818 to about 10 base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region;
- 15 provided that at least one of said second, third, and fourth regions is from Ad6.

1 MAPITAYSQQ TRGLLCIIT SLTGRDKNQV EGEVQVVSTA TQSFLATCVN
51 GVCWTVYHGA GSCTLAGPKG PITQMYTNVD QDLVGWQAPP GARSLTPCTC
101 GSSDLYLVTR HADVIPVRRR GDSRGSSLSP RPVSYLGSS GGPLLCPGSH
151 AVGIFRAAVC TRGVAKAVDF VPVESMETTM RSPVFTDNSS PPAVPQSFQV
201 AHLHAPTGSG KSTKVPAAYA AQGYKVLVLN PSVAATLGFG AYMSKAHGID
251 PNIRTGVRTI TTGAPVTYST YGKFLADGGC SGGAYDIIIC DECHSTDSTT
301 ILGIGTVLDQ AETAGARLVV LATATPPGSV TVPHPNIEEV ALSNTGEIPF
351 YGKAIPIEAI RGGRHLIFCH SKKKCDELAA KLSGLGINAV AYYRGLDVSV
401 IPTIGDVVVV ATDALMTGYT GDFDSVIDCN TCVTQTVDFS LDPTFTIETT
451 TVPQDAVSRS QRRGRTGRGR RGIYRFVTPG ERPSGMFDSS VLCECYDAGC
501 AWYELTPAET SVRLRAYLNT PGLPVCQDHIL EFWESVFTGL THIDAHFLSQ
551 TKQAGDNFPY LVAYQATVCA RAQAPPPSWD QMWKCLIRLK PTLHGPTPLL
601 YRLGAVQNEV TLTHPITKYI MACMSADLEV VTSTWVLVGG VLAALAAYCL
651 TTGSVVIVGR IILSGRPAIV PDREFLYQEF DEMEECASHL PYIEQGMQLA
701 EQFKQKALGL LQTATKQAEA AAPVVESKWR ALETFWAKHM WNFISGIQYL
751 AGLSTLPGNP AIASLMAFTA SITSPLTTQS TLLFNILGGW VAAQLAPPSA
801 ASAFCVGAGIA GAAVGSIQLG KVLDILAGY GAGVAGALVA FKVMMSGEMPS
851 TEDLVNLLPA ILSPGALVVG VVCAAIRRH VGPGEHAVQW MNRLIAFASR
901 GNHVSPTHYV PESDAAARVT QILSSLTITQ LLKRLHQWIN EDCSTPCSGS
951 WLRDVWDWIC TVLTDFTKWL QSKLLPQLPG VPFFSCQRGY KGVWRGDGIM
1001 QTTCPGQAQI TGHVKNGSMR IVGPKTCSNT WHGTFPINAY TTGPCTPSA
1051 PNYSRALWRV AAEEYVEVTR VGDFHYVTGM TTDNVKCPCQ VPAPEFFTEV
1101 DGVRLHRYAP ACRPLLREEV TFQVGLNQYL VGSQLPCEPE PDVALTSML
1151 TDPSHITAET AKRRLARGSP PSLASSSSASQ LSAPSLKATC TTHHVSPDAD
1201 LIEANLLWRQ EMGGNITRVE SENKVVVLDs FDPLRAEEDe REVSPAEIL
1251 RKSKKFPAAM PIWARPDYNP PLLESWKDPD YVPPVVHGCP LPPIKAPPiP
1301 PPRRKRTVVL TESSVSSALA ELATKTFGSS ESSAVDSGTA TALPDQASDD
1351 GDKGSDVESy SSMPPLEGEP GDPDLSDGSw STVSEEASEd VVCCSMSYTw
1401 TGALITPCAa EESKLPINAL SNSLLRHNM VYATTSRSAg LRQKKVTFDR
1451 LQVLDDHYRD VLKEMKAKAS TVKAKLLSVE EACKLTPPHs AKSKFGYGAk
1501 DVRNLSSKAV NHIHSVWKDL LEDTVTPIDT TIMAKNEVFC VQPEKGGRKP
1551 ARLIVFPDLG VRVCEKMALY DVVSTLPQVV MGSSYGFQYS PGQRVEFLVN
1601 TWKSKKNPmg FSYDTRCFDS TVTENDIRVE ESIYQCCDLA PEARQAiKSL
1651 TERLYIGGPL TNSKGQNCGY RRCRASGVLT TSCGNLTcy LKASAACRAA

2/92

1701 KLQDCTMLVN AAGLVVICES AGTQEDAASL RVFTEAMTRY SAPPGDPPQP
1751 EYDLELITSC SSNVSVAHDA SGKRVYYLTR DPTTPLARAA WETARHTPVN
1801 SWLGNIIIMYA PTLWARMILM THFFSILLAQ EQLEKALDCQ IYGACYSIEP
1851 LDLPQIIERL HGLSAFSLHS YSPGEINRVA SCLRKLGVPP LRVWRHRARS
1901 VRARLLSQGG RAATCGKYLF NWAVTKLKL TPIPAASQLD LSGWFVAGYS
1951 GGDIYHSLSR ARPRWFMLCL LLLSVGVGIIY LLPNR

FIG. 1B

3/92

1 GCCACCATGG CGCCCATCAC GGCCTACTCC CAACAGACGC GGGGCCTACT
51 TGGTTGCATC ATCACTAGCC TTACAGGCCG GGACAAGAAC CAGGTCGAGG
101 GAGAGGTTCA GGTGGTTCC ACCGCAACAC AATCCTTCCT GGCGACCTGC
151 GTCAACGGCG TGTGTTGGAC CGTTTACCAT GGTGCTGGCT CAAAGACCTT
201 AGCCGGCCA AAGGGGCCAA TCACCCAGAT GTACACTAAT GTGGACCAGG
251 ACCTCGTCGG CTGGCAGGCG CCCCCCGGGG CGCGTCCCTT GACACCATGC
301 ACCTGTGGCA GCTCAGACCT TTACTTGTC ACAGAGACATG CTGACGTCAT
351 TCCGGTGCAC CGGCGGGCG ACAGTAGGGG GAGCCTGCTC TCCCCCAGGC
401 CTGTCTCCTA CTTGAAGGGC TCTTCGGGTG GTCCACTGCT CTGCCCTTCG
451 GGGCACGCTG TGGGCATCTT CCGGGCTGCC GTATGCACCC GGGGGGTTGC
501 GAAGGCGGTG GACTTTGTGC CCGTAGAGTC CATGGAAACT ACTATGCGGT
551 CTCCGGTCTT CACGGACAAC TCATCCCCC CGGCCGTACC GCAGTCATTT
601 CAAGTGGCCC ACCTACACGC TCCCACGGC AGCGGCAAGA GTACTAAAGT
651 GCCGGCTGCA TATGCAGCCC AAGGGTACAA GGTGCTCGTC CTCAATCCGT
701 CCGTTGCCGC TACCTTAGGG TTTGGGGCGT ATATGTCTAA GGCACACGGT
751 ATTGACCCCCA ACATCAGAAC TGGGGTAAGG ACCATTACCA CAGGCGCCCC
801 CGTCACATAC TCTACCTATG GCAAGTTCT TGCCGATGGT GGTTGCTCTG
851 GGGGCGCTTA TGACATCATA ATATGTGATG AGTGCATTC AACTGACTCG
901 ACTACAATCT TGGGCATCGG CACAGTCCTG GACCAAGCGG AGACGGCTGG
951 AGCGCGGCTT GTCGTGCTCG CCACCGCTAC GCCTCCGGGA TCGGTACCCG
1001 TGCCACACCC AAACATCGAG GAGGTGGCCC TGTCTAATAC TGGAGAGATC
1051 CCCTTCTATG GCAAAGCCAT CCCCATTGAA GCCATCAGGG GGGGAAGGCA
1101 TCTCATTTTC TGTCATTCCA AGAAGAAGTG CGACGAGCTC GCGCAAAGC
1151 TGTCAAGGCCT CGGAATCAAC GCTGTGGCGT ATTACCGGGG GCTCGATGTG
1201 TCCGTACATAC CAACTATCGG AGACGTCGTT GTCGTGGCAA CAGACGCTCT
1251 GATGACGGGC TATAACGGCG ACTTTGACTC AGTGATCGAC TGTAACACAT
1301 GTGTCACCCA GACAGTCGAC TTCAGCTTGG ATCCCACCTT CACCATTGAG
1351 ACGACGACCG TGCCCTCAAGA CGCAGTGTGG CGCTCGCAGC GGCGGGGTAG
1401 GACTGGCAGG GGTAGGAGAG GCATCTACAG GTTTGTGACT CGGGGAGAAC
1451 GGCCCTCGGG CATGTTCGAT TCCTCGGTCC TGTGTGAGTG CTATGACGCG
1501 GGCTGTGCTT GGTACGAGCT CACCCCCGCC GAGACCTCGG TTAGGTTGCG
1551 GGCCTACCTG AACACACCAG GGTGCCCCGT TTGCCAGGAC CACCTGGAGT
1601 TCTGGGAGAG TGTCTTCACA GGCCCTCACCC ACATAGATGC ACACCTCTTG
1651 TCCCAAGACCA AGCAGGCAGG AGACAACCTTC CCCTACCTGG TAGCATACCA

FIG. 2A

4/92

1701 AGCCACGGTG TGCGCCAGGG CTCAGGCCAC ACCTCCATCA TGGGATCAAA
1751 TGTGGAAGTG TCTCATACGG CTGAAACCTA CGCTGCACGG GCCAACACCC
1801 TTGCTGTACA GGCTGGGAGC CGTCCAAAAT GAGGTACCCC TCACCCACCC
1851 CATAACCAAA TACATCATGG CATGCATGTC GGCTGACCTG GAGGTCGTCA
1901 CTAGCACCTG GGTGCTGGTG GGCGGAGTCC TTGCAGCTCT GGCGCGTAT
1951 TGCCTGACAA CAGGCAGTGT GGTCAATTGTG GGTAGGATTA TCTTGTCCGG
2001 GAGGCCGGCT ATTGTTCCCG ACAGGGAGTT TCTCTACCAG GAGTTCGATG
2051 AAATGGAAGA GTGCGCCTCG CACCTCCCTT ACATCGAGCA GGGAAATGCAG
2101 CTCGCCGAGC AATTCAAGCA GAAAGCGCTC GGGTTACTGC AAACAGCCAC
2151 CAAACAAAGCG GAGGCTGCTG CTCCCGTGGT GGAGTCCAAG TGGCGAGCCC
2201 TTGAGACATT CTGGGCGAAG CACATGTGGA ATTTCATCAG CGGGATAACAG
2251 TACTTAGCAG GCTTATCCAC TCTGCCTGGG AACCCCGCAA TAGCATCATT
2301 GATGGCATTAC ACAGCCTCTA TCACCAGCCC GCTCACCACC CAAAGTACCC
2351 TCCTGTTAA CATCTGGGG GGGTGGGTGG CTGCCCAACT CGCCCCCCCC
2401 AGCGCCGCTT CGGCTTCGT GGGCGCCGGC ATCGCCGGTG CGGCTGTTGG
2451 CAGCATAGGC CTTGGGAAGG TGCTTGTGGA CATTCTGGCG GGTTATGGAG
2501 CAGGAGTGGC CGGCGCGCTC GTGGCCTTCA AGGTCAATGAG CGGCGAGATG
2551 CCCTCCACCG AGGACCTGGT CAATCTACTT CCTGCCATCC TCTCTCCTGG
2601 CGCCCTGGTC GTCGGGGTCG TGTGTGCAGC AATACTGCGT CGACACGTGG
2651 GTCCGGGAGA GGGGGCTGTG CAGTGGATGA ACCGGCTGAT AGCGTTCGCC
2701 TCGCGGGGTA ATCATGTTTC CCCCACGCAC TATGTGCCTG AGAGCGACGC
2751 CGCAGCGCGT GTTACTCAGA TCCTCTCCAG CCTTACCATC ACTCAGCTGC
2801 TGAAAAGGCT CCACCAGTGG ATTAATGAAG ACTGCTCCAC ACCGTGTTCC
2851 GGCTCGTGGC TAAGGGATGT TTGGGACTGG ATATGCACGG TGTTGACTGA
2901 CTTCAAGACC TGGCTCCAGT CCAAGCTCCT GCCGCAGCTA CCGGGAGTCC
2951 CTTTTTCTC GTGCCAACGC GGGTACAAGG GAGTCTGGCG GGGAGACGGC
3001 ATCATGAAA CCACCTGCCAC ATGTGGAGCA CAGATCACCG GACATGTCAA
3051 AAACGGTTCC ATGAGGATCG TCGGGCCTAA GACCTGCAGC AACACGTGGC
3101 ATGGAACATT CCCCACCAAC GCATACACCA CGGGCCCCCTG CACACCCTCT
3151 CCAGCGCCAA ACTATTCTAG GGCCTGTGG CGGGTGGCCG CTGAGGAGTA
3201 CGTGGAGGTC ACGCGGGTGG GGGATTTCCA CTACGTGACG GGCATGACCA
3251 CTGACAACGT AAAGTGCCCA TGCCAGGTT CCGCTCCTGA ATTCTTCACG
3301 GAGGTGGACG GAGTGCAGGTT GCACAGGTAC GCTCCGGCGT GCAGGCCTCT
3351 CCTACGGGAG GAGGTTACAT TCCAGGTCGG GCTCAACCAA TACCTGGTTG

FIG. 2B

5/92

3401 GGTACACAGCT ACCATGCGAG CCCGAACCGG ATGTAGCAGT GCTCACTTCC
 3451 ATGCTCACCG ACCCCCTCCC CATCACAGCA GAAACGGCTA AGCGTAGGTT
 3501 GGCCAGGGGG TCTCCCCCT CCTTGGCCAG CTCTTCAGCT AGCCAGTTGT
 3551 CTGCGCCTTC CTTGAAGGCG ACATGCACTA CCCACCAGTG CTCTCCGGAC
 3601 GCTGACCTCA TCGAGGCCAA CCTCCTGTGG CGGCAGGAGA TGGGCGGGAA
 3651 CATCACCCGC GTGGAGTCGG AGAACAAAGGT GGTAGTCCTG GACTCTTCG
 3701 ACCCGCTTCG AGCGGAGGGAG GATGAGAGGG AAGTATCCGT TCCGGCGGAG
 3751 ATCCTGCGGA AATCCAAGAA GTTCCCCGCA GCGATGCCA TCTGGGCGCG
 3801 CCCGGATTAC AACCCCTCCAC TGTTAGAGTC CTGGAAGGAC CCGGACTACG
 3851 TCCCTCCGGT GGTGCACGGG TGCCCGTTGC CACCTATCAA GGCCCTCCA
 3901 ATACCACCTC CACGGAGAAA GAGGACGGTT GTCCTAACAG AGTCCTCCGT
 3951 GTCTTCTGCC TTAGCGGAGC TCGCTACTAA GACCTTCGGC AGCTCCGAAT
 4001 CATCGGCCGT CGACAGCGC ACGGCGACCG CCCTTCCTGA CCAGGCCTCC
 4051 GACGACGGTG ACAAAAGGATC CGACGTTGAG TCGTACTCCT CCATGCC
 4101 CCTTGAGGGG GAACCGGGGG ACCCCGATCT CAGTGACGGG TCTTGGTCTA
 4151 CCGTGAGCGA GGAAGCTAGT GAGGATGTG TCTGCTGCTC AATGTCTAC
 4201 ACATGGACAG GCGCCTGAT CACGCCATGC GCTGCGGAGG AAAGCAAGCT
 4251 GCCCATCAAC GCGTTGAGCA ACTCTTGCT GCGCCACCAT AACATGGTTT
 4301 ATGCCACAAC ATCTCGCAGC GCAGGCCTGC GGCAGAAGAA GGTCACCTT
 4351 GACAGACTGC AAGTCCTGGA CGACCACTAC CGGGACGTGC TCAAGGAGAT
 4401 GAAGGCGAAG GCGTCCACAG TTAAGGCTAA ACTCCTATCC GTAGAGGAAG
 4451 CCTGCAAGCT GACGCCCGCA CATTGGCCA AATCCAAGTT TGGCTATGGG
 4501 GCAAAGGACG TCCGGAACCT ATCCAGCAAG GCCGTTAACC ACATCCACTC
 4551 CGTGTGGAAG GACTTGTGG AAGACACTGT GACACCAATT GACACCACCA
 4601 TCATGGAAA AAATGAGGTT TTCTGTGTCC AACCAAGAGAA AGGAGGCCGT
 4651 AAGCCAGCCC GCCTTATCGT ATTCCAGAT CTGGGAGTCC GTGTATGCGA
 4701 GAAGATGGCC CTCTATGATG TGGTCTCCAC CCTTCCTCAG GTCGTGATGG
 4751 GCTCCTCATA CGGATTCCAG TACTCTCCTG GGCAGCGAGT CGAGTTCTG
 4801 GTGAATACCT GGAAATCAA AAAAACCCC ATGGGTTTT CATATGACAC
 4851 TCGCTGTTTC GACTAACCG TCACCGAGAA CGACATCCGT GTTGAGGAGT
 4901 CAATTACCA ATGTTGTGAC TTGGCCCCCG AAGCCAGACA GGCCATAAAA
 4951 TCGCTCACAG AGCGGCTTTA TATCGGGGGT CCTCTGACTA ATTCAAAAGG
 5001 GCAGAACTGCA GGTATCGCC GGTGCCGCGC GAGCGGCGTG CTGACGACTA
 5051 GCTGCGGTAA CACCCCTCAC A TGTTACTTGA AGGCCTCTGC AGCCTGTCGA

FIG. 2C

6/92

5101 GCTGCGAAGC TCCAGGACTG CACGATGCTC GTGAACGCCG CCGGCCTTGT
5151 CGTTATCTGT GAAAGCGCGG GAACCCAAGA GGACGCGCG AGCCTACGAG
5201 TCTTCACCGA GGCTATGACT AGGTACTCTG CCCCCCCCCGG GGACCCGCC
5251 CAACCAGAAC ACGACTTGGA GCTGATAACA TCATGTTCCCT CCAATGTGTC
5301 GGTCGCCAC GATGCATCAG GCAAAAGGGT GTACTACCTC ACCCGTGATC
5351 CCACCAACCC CCTCGCACGG GCTCGTGGG AAACAGCTAG ACACACTCCA
5401 GTTAACTCCT GGCTAGGCAA CATTATCATG TATGCCCCA CTTTGTGGC
5451 AAGGATGATT CTGATGACTC ACTTCTTCTC CATCCTCTA GCACAGGAGC
5501 AACTTGAAAA AGCCCTGGAC TGCCAGATCT ACGGGGCCTG TTACTCCATT
5551 GAGCCACTTG ACCTACCTCA GATCATTGAA CGACTCCATG GCCTTAGCGC
5601 ATTTCACTC CATACTTACT CTCCAGGTGA GATCAATAGG GTGGCTTCAT
5651 GCCTCAGGAA ACTTGGGTA CCACCCCTGC GAGTCTGGAG ACATCGGGCC
5701 AGGAGCGTCC GCGCTAGGCT ACTGTCCCAG GGGGGGAGGG CCGCCACTTG
5751 TGGCAAGTAC CTCTTCAACT GGGCAGTGAA GACCAAACTC AACTCACTC
5801 CAATCCCAGC TGCGTCCCAG CTGGACTTGT CCGGCTGGTT CGTTGCTGGT
5851 TACAGCGGGG GAGACATATA TCACAGCCTG TCTCGTCCCC GACCCCGCTG
5901 GTTCATGCTG TGCCTACTCC TACTTTCTGT AGGGTAGGC ATCTACCTGC
5951 TCCCCAACCG ATAAA

FIG. 2D

7/92

1 GCCACCATGG CCCCATCAC CGCCTACAGC CAGCAGACCC GCGGCCTGCT
 51 GGGCTGCATC ATCACCAAGCC TGACCGGCCG CGACAAGAAC CAGGTGGAGG
 101 GCGAGGTGCA GGTGGTGAGC ACCGCCACCC AGAGCTTCCT GGCCACCTGC
 151 GTGAACGGCG TGTGCTGGAC CGTGTACCAAC GGCGCCGGCA GCAAGACCCCT
 201 GGCCGGCCCC AAGGGCCCCA TCACCCAGAT GTACACCAAC GTGGACCAGG
 251 ACCTGGTGGG CTGGCAGGCC CCCCGCCGGG CCCGCAGCCT GACCCCCCTGC
 301 ACCTGCGGCA GCAGCGACCT GTACCTGGTG ACCCGCCACG CCGACGTGAT
 351 CCCCGTGCGC CGCCGCGGCG ACAGCCGCGG CAGCCTGCTG AGCCCCCGCC
 401 CCGTGAGCTA CCTGAAGGGC AGCAGCGGCCG GCCCCCTGCT GTGCCCCAGC
 451 GGCACGCCG TGGGCATCTT CCGCGCCGCC GTGTGCACCC GCGGCCTGGC
 501 CAAGGCCGTG GACTTCGTGC CCGTGGAGAG CATGGAGACC ACCATGCGCA
 551 GCCCCGTGTT CACCGACAAC AGCAGCCCCC CCGCCGTGCC CCAGAGCTTC
 601 CAGGTGGCCC ACCTGCACGC CCCCACCGGC AGCGGCAAGA GCACCAAGGT
 651 GCCCCGCCG TACGCCGCC AGGGCTACAA GGTGCTGGTG CTGAACCCCA
 701 GCGTGGCCGC CACCCCTGGC TTCGGCGCCT ACATGAGCAA GGCCCACGGC
 751 ATCGACCCCCA ACATCCGCAC CGCGGTGCGC ACCATCACCA CGGGCGCCCC
 801 CGTGACCTAC AGCACCTACG GCAAGTTCTT GGCGACGGC GGCTGCAGCG
 851 GCGGCGCCTA CGACATCATC ATCTGCGACG AGTGCCACAG CACCGACAGC
 901 ACCACCATCC TGGGCATCGG CACCGTGCTG GACCAGGCCG AGACCGCCGG
 951 CGCCCGCCTG GTGGTGTGG CCACCGCCAC CCCCGCCGGC AGCGTGACCG
 1001 TGCCCCACCC CAACATCGAG GAGGTGGCCC TGAGCAACAC CGGCGAGATC
 1051 CCCTTCTACG GCAAGGCCAT CCCCATCGAG GCCATCCGCG GCGGCCGCCA
 1101 CCTGATCTTC TGCCACAGCA AGAAGAAGTG CGACGAGCTG GCGGCCAAGC
 1151 TGAGCGGCCCT GGGCATCAAC GCCGTGGCCT ACTACCGCGG CCTGGACGTG
 1201 AGCGTGTATCC CCACCATCGG CGACGTGGTG GTGGTGGCCA CCGACGCCCT
 1251 GATGACCGGC TACACCGCG ACTTCGACAG CGTGATCGAC TGCAACACCT
 1301 GCGTGACCCA GACCGTGGAC TTCAGCCTGG ACCCCACCTT CACCATCGAG
 1351 ACCACCAACG TGCCCCAGGA CGCCGTGAGC CGCAGCCAGC GCGCGGGCCG
 1401 CACCGGCCGC GGCCGCCCGC GCATCTACCG CTTCGTGACC CCCGGCGAGC
 1451 GCCCCAGCGG CATGTTGAC AGCAGCGTGC TGTGCGAGTG CTACGACGCC
 1501 GGCTGCGCCT GGTACGAGCT GACCCCCGCC GAGACCAGCG TGCGCCTGGC
 1551 CGCCTACCTG AACACCCCCG GCCTGCCGT GTGCCAGGAC CACCTGGAGT
 1601 TCTGGGAGAG CGTGTTCACC GGCTGACCC ACATCGACGC CCACCTCCTG
 1651 AGCCAGACCA AGCAGGCCGG CGACAACTTC CCCTACCTGG TGGCCTACCA

FIG. 3A

8/92

1701 GGCCACCGTG TGCGCCCGCG CCCAGGCCCC CCCCCCCCAGC TGGGACCAGA
 1751 TGTGGAAGTG CCTGATCCGC CTGAAGCCCA CCCTGCACGG CCCCACCCCC
 1801 CTGCTGTACC GCCTGGCGC CGTGCAGAAC GAGGTGACCC TGACCCACCC
 1851 CATCACCAAG TACATCATGG CCTGCATGAG CGCCGACCTG GAGGTGGTGA
 1901 CCAGCACCTG GGTGCTGGTG GGCGGCGTGC TGGCCGCCCT GGCCGCCTAC
 1951 TGCCTGACCA CGGGCAGCGT GGTGATCGTG GGCCGCATCA TCCTGAGCGG
 2001 CCGCCCCGCC ATCGTGCCCG ACCCGAGTT CCTGTACCAAG GAGTTCGACG
 2051 AGATGGAGGA GTGCGCCAGC CACCTGCCCT ACATCGAGCA GGGCATGCAG
 2101 CTGGCCGAGC AGTTCAAGCA GAAGGCCCTG GGCCTGCTGC AGACCGCCAC
 2151 CAAGCAGGCC GAGGCCGCCG CCCCCGTGGT GGAGAGCAAG TGGCGCGCCC
 2201 TGGAGACCTT CTGGGCCAAG CACATGTGGA ACTTCATCAG CGGCATCCAG
 2251 TACCTGGCCG GCCTGAGCAC CCTGCCCGC AACCCCGCCA TCGCCAGCCT
 2301 GATGGCCTTC ACCGCCAGCA TCACCAGCCC CCTGACCACC CAGAGCACCC
 2351 TGCTGTTCAA CATCCTGGGC GGCTGGGTGG CCGCCCAGCT GGCCCCCCCC
 2401 AGCGCCGCCA GCGCCTTCGT GGGCGCCGGC ATCGCCGGCG CCGCCGTGGG
 2451 CAGCATCGGC CTGGGCAAGG TGCTGGTGGA CATCCTGGCC GGCTACGGCG
 2501 CCGCGTGGC CGGCGCCCTG GTGGCCTTCA AGGTGATGAG CGCGAGATG
 2551 CCCAGCACCG AGGACCTGGT GAACCTGCTG CCCGCCATCC TGAGCCCCGG
 2601 CGCCCTGGTG GTGGCGTGG TGTGCGCCGC CATCCTGCGC CGCCACGTGG
 2651 GCCCCGGCGA GGGCGCCGTG CAGTGGATGA ACCGCCTGAT CGCCTTCGCC
 2701 AGCCGCGGCA ACCACGTGAG CCCCACCCAC TACGTGCCCG AGAGCGACGC
 2751 CGCCGCCCGC GTGACCCAGA TCCTGAGCAG CCTGACCATC ACCCAGCTGC
 2801 TGAAGCGCCT GCACCAAGTGG ATCAACGAGG ACTGCAGCAC CCCCTGCAGC
 2851 GGCAGCTGGC TGGCGACGT GTGGGACTGG ATCTGCACCG TGCTGACCGA
 2901 CTTCAAGACC TGGCTGCAGA GCAAGCTGCT GCCCCAGCTG CCCGGCGTGC
 2951 CCTCTTCAG CTGCCAGCGC GGCTACAAGG GCGTGTGGCG CGGCGACGGC
 3001 ATCATGCAGA CCACCTGCCCT CGCGGGCGCC CAGATCACCG GCCACGTGAA
 3051 GAACGGCAGC ATGCGCATCG TGGGCCCCAA GACCTGCAGC AACACCTGGC
 3101 ACGGCACCTT CCCCCATCAAC GCCTACACCA CGGGCCCCCTG CACCCCCAGC
 3151 CCGCCCCCCTA ACTACAGCCG CGCCCTGTGG CGCGTGGCCG CCGAGGAGTA
 3201 CGTGGAGGTG ACCCGCGTGG GCGACTTCCA CTACGTGACC GGCATGACCA
 3251 CCGACAACGT GAAGTGCCCG TGCCAGGTGC CCGCCCCCGA GTTCTTCACC
 3301 GAGGTGGACG GCGTGCCTGCACCGCTAC GCCCCCGCCT GCCGCCCCCT
 3351 GCTGCACCGAG GAGGTGACCT TCCAGGTGGG CCTGAACCAG TACCTGGTGG

FIG. 3B

9/92

3401 GCAGCCAGCT GCCCTGCGAG CCCGAGCCCC ACGTGGCCGT GCTGACCAGC
 3451 ATGCTGACCG ACCCCAGCCA CATCACCGCC GAGACGCCA AGCGCCGCCT
 3501 GGCCCAGCGC AGCCCCCCC GAACCGCCAG CAGCAGCGCC AGCCAGCTGA
 3551 GCGCCCCCAG CCTGAAGGCC ACCTGCACCA CCCACACGT GAGCCCCGAC
 3601 GCCGACCTGA TCGAGGCCAA CCTGCTGTGG CGCCAGGAGA TGGGCGGCAA
 3651 CATCACCCGC GTGGAGAGCG AGAACAAAGGT GGTGGTGCTG GACAGCTTCG
 3701 ACCCCCTGCG CGCCGAGGAG GACGAGCGCG AGGTGAGCGT GCCCGCCGAG
 3751 ATCCTGCGCA AGAGCAAGAA GTTCCCCGCC GCCATGCCA TCTGGGCCCG
 3801 CCCCCACTAC AACCCCCCCC TGCTGGAGAG CTGGAAGGAC CCCGACTACG
 3851 TGCCCCCGT GGTGCACGGC TGCCCCCTGC CCCCCATCAA GGCCCCCCCC
 3901 ATCCCCCCCCC CCCGCCGCAA GCGCACCGTG GTGCTGACCG AGAGCAGCGT
 3951 GAGCAGCGCC CTGGCCGAGC TGGCCACCAA GACCTTCGGC AGCAGCGAGA
 4001 GCAGCGCCGT GGACAGCGGC ACCGCCACCG CCCTGCCGA CCAGGCCAGC
 4051 GACGACGGCG ACAAGGGCAG CGACGTGGAG AGCTACAGCA GCATGCCCG
 4101 CCTGGAGGGC GAGCCCCGGCG ACCCCGACCT GAGCGACGGC AGCTGGAGCA
 4151 CCGTGAGCGA GGAGGCCAGC GAGGACGTGG TGTGCTGCGAG CATGAGCTAC
 4201 ACCTGGACCG GCGCCCTGAT CACCCCCCTGC GCCGCCGAGG AGAGCAAGCT
 4251 GCCCATCAAC GCCCTGAGCA ACAGCCTGCT GCGCCACAC AACATGGTGT
 4301 ACGCCACAC CAGCCGCAGC GCCGGCCTGC GCCAGAAGAA GGTGACCTTC
 4351 GACCGCCTGC AGGTGCTGGA CGACCACTAC CGCGACGTGC TGAAGGAGAT
 4401 GAAGGCCAAG GCCAGCACCG TGAAGGCCAA GCTGCTGAGC GTGGAGGAGG
 4451 CCTGCAAGCT GACCCCCCCC CACAGGCCA AGAGCAAGTT CGGCTACGGC
 4501 GCCAAGGACG TGCGCAACCT GAGCAGCAAG GCCGTGAACC ACATCCACAG
 4551 CGTGTGGAAG GACCTGCTGG AGGACACCGT GACCCCCATC GACACCACCA
 4601 TCATGGCCAA GAACGAGGTG TTCTGCGTGC AGCCCGAGAA GGGCGGCCGC
 4651 AAGCCGCCGC GCCTGATCGT GTTCCCCGAC CTGGGCGTGC GCGTGTGCGA
 4701 GAAGATGGCC CTGTACGACG TGGTGAGCAC CCTGCCCGAG GTGGTGTGG
 4751 GCAGCAGCTA CGGCTTCCAG TACAGCCCCG GCCAGCGCGT GGAGTTCTG
 4801 GTGAACACCT GGAAGAGCAA GAAGAACCCC ATGGGCTTCA GCTACGACAC
 4851 CCGCTGCTTC GACAGCACCG TGACCGAGAA CGACATCCGC GTGGAGGAGA
 4901 GCATCTACCA GTGCTGCGAC CTGGCCCCCG AGGCCCCGCC GGCCATCAAG
 4951 AGCCTGACCG AGCGCCTGTA CATCGGCGGC CCCCTGACCA ACAGCAAGGG
 5001 CCAGAACTGC GGCTACCGCC GCTGCCGCGC CAGCGCGTGC CTGACCACCA
 5051 GCTGCGGCAA CACCCCTGACC TGCTACCTGA AGGCCAGCGC CGCCTGCCGC

FIG. 3C

10/92

5101 GCCGCCAAGC TGCAGGACTG CACCATGCTG GTGAACGCCG CGGGCCTGGT
5151 GGTGATCTGC GAGAGCCCG GCACCCAGGA GGACGCCGCC AGCCTGCGCG
5201 TGTTCACCGA GGCCATGACC CGCTACAGCG CCCCCCCCCCG CGACCCCCCCC
5251 CAGCCCGAGT ACGACCTGGA GCTGATCACC AGCTGCAGCA GCAACGTGAG
5301 CGTGGCCCAC GACGCCAGCG GCAAGCGCGT GTACTACCTG ACCCGCGACC
5351 CCACCACCCC CCTGGCCCGC GCCGCCTGGG AGACCGCCCG CCACACCCCC
5401 GTGAACAGCT GGCTGGCAA CATCATCATG TACGCCCCA CCCTGTGGGC
5451 CCGCATGATC CTGATGACCC ACTTCCTTCAG CATCCTGCTG GCCCAGGAGC
5501 AGCTGGAGAA GGCCCTGGAC TGCCAGATCT ACGGCGCCTG CTACAGCATC
5551 GAGCCCTGG ACCTGCCCA GATCATCGAG CGCCTGCACG GCCTGAGCGC
5601 CTTCAGCCTG CACAGCTACA GCCCCGGCGA GATCAACCGC GTGGCCAGCT
5651 GCCTGCGCAA GCTGGCGTG CCCCCCCTGC GCGTGTGGCG CCACCGCGCC
5701 CGCAGCGTGC GCGCCCGCCT GCTGAGCCAG GGCGGCCGCG CCGCCACCTG
5751 CGGCAAGTAC CTGTTCAACT GGGCCGTGAA GACCAAGCTG AAGCTGACCC
5801 CCATCCCCGC CGCCAGCCAG CTGGACCTGA GCGGCTGGTT CGTGGCCGGC
5851 TACAGCGGCG GCGACATCTA CCACAGCCTG AGCCGCGCCC GCCCCCGCTG
5901 GTTCATGCTG TGCCTGCTGC TGCTGAGCGT GGGCGTGGGC ATCTACCTGC
5951 TGCCCAACCG CTAAA

FIG. 3D

11/92

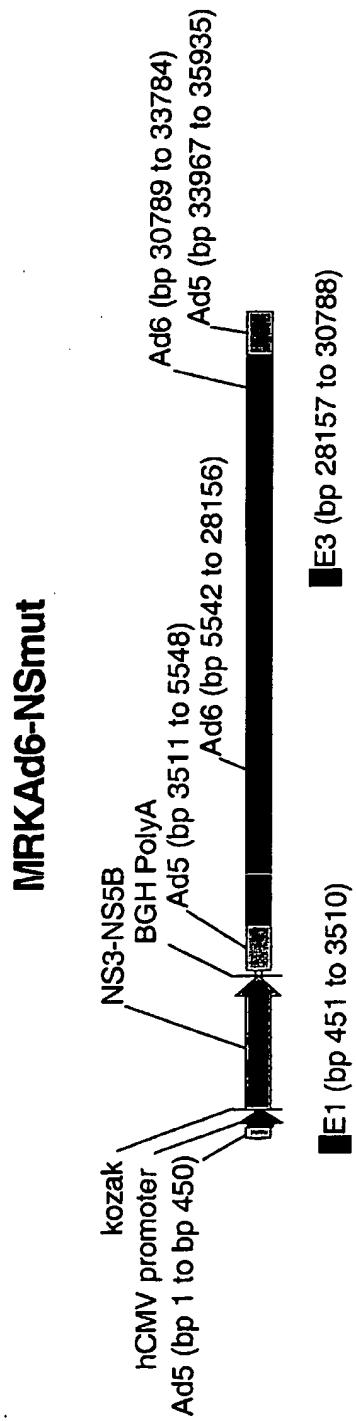


FIG. 4A

12/92

1 catcatcaat aataaacctt atttggatt gaagccaata tgataatgag ggggtggagt
 61 ttgtgacgtg gcgcggggcg tgggaacggg gcgggtgacg tagtagtgtg gcggaaagtgt
 121 gatgttgc当地 gtgtggc当地 acacatgtaa gcacggatg tggcaaaagt gacgttttg
 181 gtgtgc当地 ggtagcacacag gaagtgacaa ttttc当地 gtc当地 gatgtttag
 241 taaaatttggg cgttaaccgag taagatttg ccatttc当地 gggaaaactg aataagagga
 301 agtgaatct gaataatccc ttgttactca tagc当地 gtaa tatttgc当地 gggcc当地
 361 gactttgacc gttacgtgg agactc当地 aggtgtt当地 ctc当地 ggtt当地 ttccg当地
 421 cgggtcaaag ttggc当地 ttattatag gccccgca tccattgc当地 acgtt当地
 481 catatcataa tatgtacatt tatattggct catgtccaaac attaccgcca tggatc当地
 541 gattattgac tagtattaa tagtaatcaa ttacggggtc attagttcat agcccatata
 601 tggagttccg cgttacataa ct当地 acggtaa atggccgccc tggctgacc cccaaacgacc
 661 cccgccc当地 gacgtcaata atgacgtatg ttcccatagta aacgccaata gggacttcc
 721 attgacgtca atgggtggag tatttacggta aactgccc当地 ttggc当地 agtcaatgt
 781 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctglocal
 841 atgccc当地 agtcaata catgaccctt当地 tggactt当地 ctacttggca gtacatctac gtatttagtca
 901 tc当地 ctattac catggtgatg cgggggatc agtacatcaa tggc当地 ggtt当地 tagc当地
 961 actcacgggg atttccaaatg ctccacccca ttgacgtcaa tggagttt当地 ttttgc当地
 1021 aaaatcaacg ggactttccaaatgtaa aatgtcgta acaactccgc cccattgacg caaatggcg
 1081 gtaggc当地 gtgtgggag gtctatataa gcagagctcg tttagtgaac cgtagatcg
 1141 cctggagacg ccatccacgc tggtt当地 gacc tccatagaag acaccgggac cgatccagcc
 1201 tccgc当地 ggaacgggtc attggaacgc ggatccccg tggcaagagt gagatctg
 1261 accatggc当地 ccatcacggc ctactccaa cagacgccc gcctactt当地 ttgc当地
 1321 actagc当地 caggccgggaa caagaaccag gtc当地 gagggag aggttc当地 gaggttcc
 1381 gcaacacaat cttccctggc gacgtcgta aacggc当地 gtggacccg ttaccatgg
 1441 gctggctcaa agaccttagc cggcccaaag gggccaatca cccagatgt aactatgt
 1501 gaccaggacc tc当地 ctggc当地 gcaggccccc cccggggcgc gttccctt当地 accatgc
 1561 tgtggc当地 agtcaataa ct当地 ggacatgtc acgtc当地 tccctt当地 ggtgc当地
 1621 cggggc当地 gtagggggag cctgctcttccc cccaggc当地 tctc当地 ttttcc
 1681 tc当地 gggtc当地 cactgctctg cc当地 tccggg cacgtgtgg gcatctt当地 ggctg
 1741 tgc当地 cccggg ggggtgcaaa ggccgtggac tttt当地 gccc当地 tagatccat
 1801 atgc当地 gtcttcc cggctt当地 cacatgtcc ggacaactca tccccccgg cgttaccgca
 1861 gtggcc当地 accaccgtcc cactggc当地 ggcaagagta ctaaagtgcc ggctg
 1921 gcaagccaaag ggtacaaggt gtc当地 ctcc aatcgttcc ttggc当地 tctagg
 1981 ggggc当地 tgc当地 taaggc acacggtaa gaccatca tc当地 aactgg ggtaggacc
 2041 attaccacag ggc当地 cccgggctg cacatcttcc acctatggca agtttctt当地 cgatgt
 2101 tgctctgggg ggc当地 ttatgaa catcataata tgc当地 tttt当地 ggc当地
 2161 acaatcttgg gcatggc当地 tccggatcg gtc当地 accccatccc cattg
 2221 gtgctcgccaa cc当地 ctacgccc agagatcccc ttctatggca aagccatccc
 2281 gtggccctgt ctaatactgg catttctgt catttcaaga agaagtgc当地 cgagct
 2341 atcagggggg gaaaggcatgt catttctgtt cttccatggc当地 agtgc当地
 2401 gcaaaatgt caggcctc当地 aatcaacgt gtggc当地 tatttccatggc
 2461 gtc当地 ataccaa ctatgggaa cgtc当地 ttgtc当地 gtggc当地 acgtc当地
 2521 acggccgact ttactc当地 gatcgactgt aacacatgtc tc当地 cccaggac
 2581 agtgc当地 gtc当地 cc当地 acccttccatggc当地 cttcaagacgc
 2641 tc当地 gagccgccc ggggttaggac tggc当地 agggaggca tctacagg
 2701 ggagaacggc cctc当地 gatccatggc当地 ttgtc当地 tggc当地 gtc当地
 2761 tgtgctt当地 ggatcgactc当地 cccggccgag acctc当地 ggtt当地
 2821 acaccagggt tggccg当地 cc当地 aggaccac ctggagttt当地 gggag
 2881 ctc当地 accccaca tagatgc当地 cttctt当地 gagcca
 2941 tacctggtag cataccaagc caccgtt当地 gccaggc当地 agggcc
 3001 gatcaaatgt ggaagggtt当地 catacgcc当地 aaacctt当地
 3061 ctgtacaggc tgggagccgt cccaaaatgt gtc当地 accctt当地
 3121 atcatggcat gcatgtcc当地 tgacctt当地 gtc当地 cacta
 3181 ggaggc当地 ttgtccgggag gccggctattt当地 gttccg
 3241 aggattatct tggccgggag gccggctattt当地 cttt当地
 3301 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3361 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3421 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3481 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3541 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3601 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3661 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3721 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3781 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3841 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3901 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 3961 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4021 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4081 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4141 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4201 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4261 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4321 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4381 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4441 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4501 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4561 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4621 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4681 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4741 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4801 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4861 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4921 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 4981 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5041 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5101 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5161 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5221 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5281 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5341 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5401 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5461 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5521 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5581 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5641 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5701 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5761 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5821 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5881 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 5941 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6001 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6061 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6121 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6181 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6241 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6301 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6361 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6421 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6481 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6541 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6601 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6661 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6721 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6781 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6841 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6901 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 6961 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7021 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7081 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7141 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7201 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7261 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7321 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7381 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7441 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7501 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7561 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7621 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7681 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7741 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7801 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7861 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7921 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 7981 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8041 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8101 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8161 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8221 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8281 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8341 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8401 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8461 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8521 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8581 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8641 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8701 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8761 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8821 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8881 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 8941 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9001 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9061 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9121 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9181 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9241 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9301 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9361 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9421 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9481 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9541 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9601 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9661 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9721 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9781 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9841 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9901 gatgtt当地 tggccgggag gccggctattt当地 cttt当地
 9961 gatgtt当地 tggccgggag gccggctattt当地 cttt当地

FIG. 4B

13/92

3361 gccgagcaat tcaaggcagaa agcgctcggg ttactgc当地 cagccaccaa acaagcggag
3421 gctgtgtctt ccgtgggtgg aatccaaatgg cgagcccttg agacattctg ggcgaagcac
3481 atgtgaatt tcatcagcgg gatacagttac tttagcaggct tatccactct gcctggaaac
3541 cccgcaatag catcattgtat ggcattcaca gcctctatca ccagcccgt caccacccaa
3601 agtacccctcc ttgttaacat ctgggggggg tgggtggctg cccaaactcgc cccccccagc
3661 gccgcttcgg ctgttgcggg cgccggcata gcccgtgcgg ctgttgcgg cataggcctt
3721 gggaaagggtgc ttgtggacat tctggcgggt tatggagcag gagtggccgg cgccgtcgt
3781 gccttcaagg tcatcagcgg cgagatgccc tccaccggagg acctggtaa tctacttcct
3841 gccatcctct ctccctggcgc cctgtgtcggc ggggtggctgt gtgcagcaat actgcgtcga
3901 cacgtgggtc cgggagaggg ggctgtgcag tggatgaacc ggctgtatgc gttcgccctcg
3961 cggggtaatc atgtttcccc cacgcaactat gtgcgtgaga ggcacgcgc agcgcgtgtt
4021 actcagatcc tctccagcct taccatcaact cagtcgtcga aaaggctcca ccagtggatt
4081 aatgaagact gctccacacc gtgttccggc tcgtggctaa gggatgtttg ggactggata
4141 tgcacccgtgt tgactgactt caagacatttgg ctccatgtca agctcctgccc gcagctaccg
4201 ggagtccctt ttttctcggt ccaacgcggg tacaaggag tctggcgggg agacggcattc
4261 atgcaaacca cctgcccatttggagcag atcaccggac atgtcaaaaa cggtccatg
4321 aggatcgtcg ggcctaagac ctgcagcaac acgtggcatg gaacattttcc catcaacgc
4381 tacaccacgg gcccctgcac accctcttca ggcaccaacttccatg gctgtggcgg
4441 gtggccgtcg aggaggatcg ggaggtcaccg cgggtggggg atttccatca cgtgacggc
4501 atgaccactg acaacgtaaa gtgccttcaatcaccggatcgttccgg ctccatgatttccatg
4561 gtggacggag tgggttgcg caggtacgttccggcgtcga ggcctcttccatg acggaggag
4621 gttacattcc aggtcggttca accacaatac ctgttgggttccatg cagactacc atgcgagccc
4681 gaacccggatg tagcagtgttca cacttccatg ctacccggacc cctccatcatc acacccgg
4741 acggctaagc gttaggttgcg caggggttcttccatg tggccatgtc ttcatgtac
4801 cagttgtcttgcgccttccatg tgggttgcg caggtacgttccggcgtcga ggcctcttccatg acggaggag
4861 gacccatcg aggtcggttca accacaatac ctgttgggttccatg cagactacc atgcgagccc
4921 gagtcggaga acaagggtggt agtcttgcgtc tctttcgacc cgttgcgtc ggaggaggat
4981 gagaggaaatg tattccatgtc ggcggagatc ctgcggaaat ccaagaatgttccatg cccgcagcg
5041 atgcccatttccatg tggcgccccc ggattacaaatc ccttcactgt tagagtcttgcgtc tccgacgc
5101 gactacgtcc ctccgggtggt gcacgggtgc ccgttgcacc ctatcaaggc cccttcaata
5161 ccacccatcg ggagaaatgg gacgggttgc ctaacatgttccatg tggccatgtc ttctgcctt
5221 gcgagatcg ctactaagac ctccggcgcgc tccatgttccatg cggccgtcga cagccgc
5281 ggcaccggcc ttccatgttccatg ggcctccatg gacgggttccatg aaggatccga cgttgcgt
5341 tactccatcg tgccccccctt tgagggggaa ccggggggacc cccatctcgtc tgacgggttccatg
5401 tgggttccatcg tgagcgagga agcttagtgcg gatgtgttccatg gctgtcaat gtcctacaca
5461 tggacaggcg ctttgcgttccatg gccatgttccatg gcgaggaaa gcaagctgc catcaacgc
5521 tttagtgcgttccatg ccaccatgttccatg atgtttatgttccatg ccacaacatc tcgcacgc
5581 ggcctgcggc agaagaaggatcg caccatgttccatg agactgcgac ccactaccgg
5641 gacgtgttccatg aggaggatcg ggcgaaggcg tccatgttccatg aggctaaact cctatccgt
5701 gagggaaatgg tattccatgttccatg gccatgttccatg tggccatgttccatg ccaagtttgg
5761 aaggacgtcc ggaacccatcg ctttgcgttccatg gacgggttccatg tccactccgt gtggaggac
5821 ttgtgttccatg acactgttccatg accaattgttccatg accaattgttccatg tggcaaaaaa
5881 ttttgttccatg ctttgcgttccatg aggtgggttccatg ccacccggcc ttatcgatgttccatg
5941 ggaggatcg ttttgttccatg gatggccctt ttttgttccatg ttttgttccatg ttttgttccatg
6001 gtgtgttccatg ttttgttccatg attccatgttccatg ttttgttccatg ttttgttccatg
6061 aataccatgttccatg aatcaaaatgg aaacccatgttccatg ggctttcat atgacactcg
6121 tcaacccatgttccatg ccacccggcc ttttgttccatg ttttgttccatg ttttgttccatg
6181 gcccggccatgttccatg ccacccggcc ttttgttccatg ttttgttccatg ttttgttccatg
6241 ctgtacttccatg ctttgcgttccatg gacgtgttccatg ttttgttccatg ttttgttccatg
6301 acgactatgttccatg gggtaacac ctttgcgttccatg ttttgttccatg ttttgttccatg
6361 gcgaaatgttccatg aggactgttccatg gatgtgttccatg aacggccggcc gcttgcgttccatg
6421 agcggggaaatgg cccaaatgttccatg ctttgcgttccatg ttttgttccatg ttttgttccatg
6481 tactccatgttccatg cccggccatgttccatg ccacccggcc ttttgttccatg ttttgttccatg
6541 ttttgttccatg atgtgttccatg ctttgcgttccatg gatgtgttccatg ttttgttccatg
6601 ctttgcgttccatg ccacccggcc ttttgttccatg ttttgttccatg ttttgttccatg
6661 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
6721 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
6781 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
6841 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
6901 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
6961 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7021 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7081 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7141 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7201 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7261 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7321 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7381 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7441 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7501 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7561 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7621 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7681 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7741 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7801 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7861 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7921 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
7981 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8041 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8101 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8161 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8221 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8281 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8341 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8401 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8461 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8521 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8581 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8641 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8701 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8761 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8821 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8881 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
8941 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9001 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9061 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9121 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9181 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9241 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9301 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9361 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9421 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9481 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9541 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9601 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9661 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9721 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9781 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9841 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9901 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg
9961 ctttgcgttccatg ttttgttccatg ttttgttccatg ttttgttccatg ttttgttccatg

FIG. 4C

14/92

6661 aactcctggc taggcaacat tatcatgtat gcgcacctt tgtggcaag gatgattctg
 6721 atgactcaact tcttccat cttcttagca caggagcaac ttgaaaaggc cctggactgc
 6781 cagatctacg gggctgtta ctccattgag ccacttgacc tacctcgat cattgaacga
 6841 ctccatggcc ttagcgcatt ttcaactccat agttactctc caggtgagat caatagggtg
 6901 gcttcatgcc tcagggaaact tggggtagcca cccttgcag tctggagaca tcgggcccagg
 6961 agcgtccgcg ctaggctact gtccccgggg gggagggccg ccacttgtgg caagtacac
 7021 ttcaactggg cagtgaagac caaaactcaa ctcactccaa tccccgtgc gtcccagctg
 7081 gacttgtccg gctggttcg tgcgtgttac agcgggggag acatataatca cagcctgtct
 7141 cgtccccac cccgcgtgtt catgctgtgc ctactccat tttctgttagg gtaggcac
 7201 tacctgtcc ccaaccggta aatcttagac tgcgtttctt agttgcacgc catctgttgc
 7261 ttgcccctcc cccgtgcctt cttgcaccc ggaaggtgcc actcccaactg tcctttccat
 7321 ataaaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattt tgggggtgg
 7381 ggtggggcag gacagcaagg gggaggattt ggaagacaat agcaggcatg ctgggatgc
 7441 ggtgggctct atggccgatc ggcgcgcgt actggaaatgt gtgggcgtgg cttaaagggt
 7501 gaaaaagaata tataagggtgg gggtcttatg tagtttgtt tctgttttgc agcagccgc
 7561 gcccgcattga gcaccaactc gtttgcattga agcattgtga gctcatattt gacaacgcgc
 7621 atgccccccat gggccgggggt gctgtcagaat gtgatgggtt ccagcattgt tggtcgcccc
 7681 gtcctggccg caaactctac tacccgttacc tacgagaccg tgcgttggaaac gccgttggag
 7741 actgcagect cccgcgcgc ttcagccgtt gcagccaccc cccgcgggat tgcgtactgac
 7801 tttgcattcc tgagccgcgt tgcgttgcgtt gcagtttccc gtcattccgc cccgcgtatgac
 7861 aagttgacgg ctctttggc acaattggat tcttgcaccc gggaaacttaa tgcgttttct
 7921 cagcagctgt tggatctgcg ccagcagggtt tctgcctgtt aaggcttccctt ccctccaaat
 7981 gcggtttaaa acataaataaa aaaaccagac tctgtttggt tttggatcaa gcaagtgtct
 8041 tgcgttctt atttaggggt tttgcgcgcg cggttagggcc gggaccagccg gtctcggtcg
 8101 ttgagggtcc tggatgtat tttcaggacg tggtaaagggt gactctggat gttcagatac
 8161 atgggcataa gcccgtctct ggggtggagg tagcaccactt gcaagcttcc atgctcgccc
 8221 gtgggtttgtt agatgtatcca gtcgttagcag gagcgcgtgg cgtgtgttgcctt aaaaatgtct
 8281 ttcagtagca agtgcattgc caggggcgc cccctgggtt aagtgtttac aaagcggtta
 8341 agctgggatg ggtgcatacg tggggatatg agatgcattt tggactgtat ttttaggttgc
 8401 gctatgttcc cagccatc cctccgggat ttcatgttgc tgcgttgcctt gcaagaccac cagcacatgt
 8461 tatccggcgc acttggggaa tttgtcatgtt agcttagaaat gaaatgcgtt gaaacttgc
 8521 gagacgcctc tgcgttgc aagattttcc atgcatttgc cccataatgtt ggcaatgggc
 8581 ccacggccgg cggccctggc gaagatattt ctggatcac taacgtcata gttgttgc
 8641 aggatgagat cgtcataggc cattttaca aagcgcggc ggagggtgcc agactgcgtt
 8701 ataatggttc catccggccc aggggcgttgc ttacccttgc agatttgc tttccacgc
 8761 ttgagttcag atggggggat catgtctacc tgcggggcga tgaagaaaac ggtttccgg
 8821 gttaggggaga tcagctgggaa agaaagcagg ttcctgagca gctgcgactt accgcagccg
 8881 gtggggccgtt aaatcacacc tattaccgc tgcaactggt agttaagaga gctgcagctg
 8941 ccgtcatccc tgacgggggg ggcacttcg ttaagcatgt ccctgactcg catgtttcc
 9001 ctgacccaaat ccgcagaag ggcgtccgc cccagcgata gcagttctt caaggaagca
 9061 aagtttca acggtttgcg accgtccgcg gtggcatgc ttttgcgtt tgcgttgc
 9121 agtccaggc ggtcccacag ctgggttacc tgcgttgcgg catctcgatc cagcatatct
 9181 cctcggttcg cgggttgggg cggcttcgc tgcgttgc tgcgttgc tgcgttgc
 9241 gggccagggtt catgttttc cacggggcga ggggtcttgc cagcgatgc tgggttgc
 9301 tgaagggttgc cgtccgggc tgcgttgc tgcgttgc tgcgttgc
 9361 tgctgaagcg ctggccgtt tgcgttgc cgtccggccag gtgcatttgc accatgggttgc
 9421 catagtccag ccctccgcg ggcgtccgc tggcgcgc tgcgttgc tgcgttgc
 9481 cgcacgagggtt gcgttgcaga cttttgggg cgttagatgc tggcgcgc gaaatccgatt
 9541 cccggggatgtt ggcattccgcg cccgcaggccc cgcacacggc ctcgcattcc acgaggccagg
 9601 tgagctctgg cgggttgggg tcaaaaacca ggtttccccc atgcgttttgc atgcgtttcc
 9661 tacctctgtt ttccatgac cgggttccac gtcgttgc gaaaaggctt tccgttgc
 9721 cgtatacaga ctggatgggc ctgcgttgc ggcgttgc tgcgttgc
 9781 actcgatccca cttgcgttgc aaggctcgatc tccaggccag cacgaaggag gctaagtgg
 9841 aggggttagcg gtcgttgc actagggggtt ccactcgatc cagggttgc gacacatgt
 9901 cggcccttcc ggcattcaagg aaggctcgatc tccaggccag cacgaaggag gctaagtgg

FIG. 4D

9961 ttcctgaagg ggggctataa aaggggtgg gggcgcttc gtcctcactc tcttccgcatt
 10021 cgctgtctgc gagggccagc tggtgggtg agtactccct ctcaaaagcg ggcattgactt
 10081 ctgcgctaag attgtcagtt tccaaaaacg aggaggattt gatattcacc tggcccgcgg
 10141 ttagtgcctt gagggtggcc gcgtccatct ggtcagaaaa gacaatctt ttgttgtcaa
 10201 gcttgggtggc aaacgacccg tagagggcgt tggacagcaa cttggcgatg gagcgcaggg
 10261 tttgggtttt gtcgcgatcg gcgcgcctt tggccgcgt gtttagctgc acgtattcgc
 10321 gcgcaacgca ccgcattcg ggaaagacgg tggtgcgctc gtcgggact aggtgcacgc
 10381 gccaaaccgcg gtgtgcagg gtgacaaggta caacgctgtt ggctacctt ccgcgttaggc
 10441 gctcggtggc ccacgcaggc cggccgcct tgcgcgagca gaatggcggt agtgggtcta
 10501 gctgcgtctc gtccgggggg tctgcgtcca cggtaaagac cccgggcagc aggccgcgcgt
 10561 cgaagtatgc tatcttgcatt cttgcagaat ctaggcctt ctgcgcattgcg cggccggcaaa
 10621 gcgccgcctc gtatgggtt agtggggac cccatggcat ggggtgggtg agcgcggagg
 10681 cgtacatgcc gcaaatgtcg taaacgtaga gggctctt ggttatttca agatatgttag
 10741 ggttagatct tcacccgcgg atgctggcgc gcacgtatc gtatagttcg tgcgaggag
 10801 cgaggaggtc gggaccgagg ttgctacggg cggcgtgc tgcctcggaa actatctgcc
 10861 tgaagatggc atgtgagttt gatgatatgg ttgacgtgtt gaaagacgtt aagctggcg
 10921 ctgtgagacc taccgcgtca cgcacgaagg aggctgtt gtcgcgcagc ttgttgcacca
 10981 gctcgccgtt gacctgcacg tctaggcgcc agtagtccag ggtttccctt atgatgtcat
 11041 acttatccctg tccctttttt ttccacatc cgcgttgc gacaaactt tgcgggttctt
 11101 tccagactt ttggatcggaa aaccgcgttgc ccttcgcacg gtaagagccctt agcatgtaga
 11161 actgggttgcg ggccttgcgtt ggcgcacatc ctttttctt ggttgcgcg tattgcgcgc
 11221 cggccttccg gagcgggttgg tgggtgagcg caaagggttc cctaaccatg actttgaggt
 11281 actgggttgcg gaaatgcgttgc tgcgcgtatc cgccttgc tccagagcaaa aagtcgtgc
 11341 gcttttttggc acgcgggtt ggcaggcgaa aggttgcacatc gttgaagagt atcttcccc
 11401 cgcgaggcat aaatgttgcgtt gttgcgcgcg agggtcccg cacctcgaa cgggtgtt
 11461 ttacctgggc ggcgagcacg atctcgtaa agccgttgc gttgtggccc acaatgtaaa
 11521 gttccaagaa ggcggggatg cccttgcgtt aaggcaattt ttaagttcc tctgttaggt
 11581 gctcttcagg ggagctgcgc cccgtcttgc aaaggggccca gtcgcacaa tgagggttgg
 11641 aagcgcacaa tgacgtccac aggttgcgcg ccattagcat ttgcagggtgg tgcgcgaaagg
 11701 tcctaaactg ggcgcacatg gccatcccc ctgggtgtt gcaatgttgc gcaatgttgc
 11761 cttgttccca ggggtcccat ccaagggttgc cggcttagtgc tgcgcggcg gtcactatgc
 11821 gctcatctcc gcccgcactt atgaccacca tgaagggcac gagctgc tccaaaggccc
 11881 ccatccaatg ataggctctt acatcgatgg tgacaaagag acgctcggtt cgaggatgc
 11941 agccgatcg gaaagaactgg atctccccc accagttgg ggcgtggctg ttgtatgttgg
 12001 gaaagtagaa gtccttcgcg cggccgcac actcgatct gctttgtt aaacgtgcgc
 12061 agtactggca ggggtgcacg ggctgtacat ctcgcacgc gttgaccttgc gacccgc
 12121 caaggaagca ggtggaaat ttgagcccc cgcctggcg gtttggcttgg tggtctt
 12181 cttcggttgc ttgttgcgttgc cccgttgcgtt gtcgcgggg agttacgggtt gatcgacca
 12241 ccacccgcgc cggccgcac gtcgcgttgc cccgttgcgtt cggcggggatc ttgtatgc
 12301 catcgccgc atggggatgt tccatgttgc ggagctcccc cggcgttgc tcaaggccgg
 12361 gctccttcgc gtttaccttcgc catagcccc tcaggcgccg ggcttaggttcc aggttgc
 12421 tgatccatc ggggtgggtt gttggcgccgtt cgttgcgttgc caagaggccg catccccgc
 12481 ggcgcactac ggttccgcgc ggggggggtt gggccgcggg ggttgcgttgc gatgtatgc
 12541 ctaaaacggc tgacgcggc gggccccccgg aggttaggggg ggctcgccgac cccgggg
 12601 agggggcagg ggcacgttgc cggccgcgc gggcaggagc tggtgcgttgc cggcggagg
 12661 gctggcgcac ggcacgcgc ggggttgc gtcgcgttgc tggcgcctt gctgttgc
 12721 gacggccgc gtttgcgttgc accttgaaaga gagttgcaca gatcaattt cgggtgcgtt
 12781 gacggccgc tggcgttgc ttcctgcac gtcgcgttgc ttgttgcgtt gatgtatgc
 12841 ggcacatgc tgcgttgc ttccttcgc gatgttgc cgttgcgttgc gctccacgg
 12901 ggcggcgagg tgcgttgc gtcggccat gagctgcgc gggcgttgc ggcctccctc
 12961 gttcccgatcg cggctgttgc ccacgcgc ttcggcatcg cggccgcgc tgaccacctg
 13021 cgcgcgcgc ggttgcgttgc gtcggccat gagctgcgc gtcgttgc gctgttgc
 13081 gtagttgcagg tgggtggcg gtttgcgttgc cacaatgc gatataaccc agcgcgc
 13141 cgtggattcg ttgtatatccc ccaaggcctc aaggcgctt atggccttgc agaagtcc
 13201 ggcgcgcgc aaaaactggg agttgcgcgc gacacgcgtt aacttccttcc cagaagac

13261 gatgagctcg gcgacagtgt cgccgcaccc ggcgtcaaaag gctacagggg cctcttc
 13321 ttcttcaatc tcctcttcca taagggcctc ccctttct tttctggc ggggtgggg
 13381 agggggaca cggcgccac gacggcgcac cgggaggcg tcgacaaaagc gtcgatcat
 13441 ctcccccgcc cgacggcgca tggctcggt gacggcgccg ccgttctcgc gggggcgca
 13501 ttgaaagacg ccgcccgtca tggcccggtt atgggttgc ggggggctgc cgtgcggcag
 13561 ggatacggcg ctaacgtgc atctcaacaa ttgttgtga ggtactccgc caccgaggga
 13621 cctgagccag tccgcattca ccggatcgga aaacctctcg agaaaaggcg ctaaccatgc
 13681 acagtcccaa ggtaggctga gcaccgtggc gggccgcagc gggccggcggt cgggggtt
 13741 tctggccgag gtgctgtca tgatgttaaaat aaagtaggcgt gtcttgagac ggcggatgg
 13801 cgacagaagc accatgtcct tgggtccgc ctgctgaatg cgcaggcggt cggccatgcc
 13861 ccaggcttcg ttttgcacatc ggcgcaggc ttttagtag tcttgcattg gccttctac
 13921 cggcaatttc tcttctcctt cctttgtcc tgcatttcgtt gcatctatcg ctgcggcggc
 13981 ggcggagttt gggcgtaggt ggcgcctct tcctccatcg cgtgtgaccc cgaagccct
 14041 catcggtca agcaggggcca ggtcggcgac aacggcgtcg gctaataatgg cctgctgcac
 14101 ctgcgtgagg gtagactggg agtcgtccat gtccacaaaag cggtgttatg cggccgtt
 14161 gatgggttaa gtgcagggtt ccataacggc ccagttacgg gtcgtgtgac cggctgcga
 14221 gagctcggtg tacctgagac gcgagtaagc cctttagtca aagacgttagt cgttgcaagt
 14281 cccgaccagg tactgggtatc ccaccaaaaaa gtgcggcgcc ggctggcggt agaggggcca
 14341 gctgtgggtg gcccgggctc cggggccggat gtctccaaatc ataaggcgat gatatccgt
 14401 gatgtacctg gacatccagg tgatgcggc ggcgtgtggc gaggcgccg gaaagtca
 14461 gacgcgggtt cagatgttgc gcagcggcaa aaagtgcctt atggtggga cgctctggcc
 14521 ggtcaggcgcc ggcgtcgatc tgacgcgtca gaccgtgcaaa aaggagagcc tgtaagcg
 14581 cacttcccg tggctcggtg gataaattcg caagggtatc atggcgacg accgggggtt
 14641 gaaccccgga tccggccgtc cggcgtgatc catcggtta cggccggcgt gtcgaaccca
 14701 ggtgtcgac gtcagacaaac gggggaggcg tcctttggc ttccctccag ggcggccgg
 14761 tgctgcgtca gtttttttgg ccactggccg cgcggccgt aagcggttag gctggaaagc
 14821 gaaagcatttta agtggctcgc tccctgttgc cggagggtta ttttcaagg gttgagtcgc
 14881 gggacccccc gttcgagttt cggggccggcc ggactgcggc gAACGGGGGT ttgcctcccc
 14941 gtcatgcatttgc accccggctt caaattttcc cggaaacagg gacgagcccc tttttgttt
 15001 ttcccgatcatccggc tgcggcagat ggcgcggccctt cctcagcagc ggcaagagca
 15061 agagcagccg cagacatgc gggcacccctc cccttccctt acccgctcgt gaggggcaac
 15121 atcccggtt gacgcggccgg cagatggta ttacgaaccc ccgcggccggc ggacccggca
 15181 ctatggatc ttggaggagg gcgaggccctt ggcgggctta ggagcgccctt ctccgtgg
 15241 acacccaaagg gtcagctga agcgacac ggcgcggccg tacgtgcgc ggcagaacact
 15301 gtttcgcgc gcgcatggcc tgaaccgcga gcggttgcgt cgcgaggagg actttgagg
 15361 ggcgttgcgtt ggcgttgcgtt gtcgcgttgcgtt ggcgttgcgtt ggcgttgcgtt
 15421 cgacgcggcc accgggattt gtcggccgc cgcacacgtt ggcgttgcgtt acctggta
 15481 cgctgtacgag cagacgggtt accaggagat taactttcaaa aaaaacttta acaaccacgt
 15541 ggcgcacgtt gttggcgccg aggagggttc tataggactt atgcattgtt gggactttgt
 15601 aagcgcgtt gggccgtt ctttgcgtt ggcgttgcgtt ggcgttgcgtt tccttata
 15661 gcaacacggc agggacaacgg aggcatttgc ggcgttgcgtt ctttgcgtt ggcgttgcgtt
 15721 gggccgtt gtcgttgcattt tgataaacat tctgcagac atagtggtgc aggagcgcc
 15781 ttggccgtt gtcgttgcattt taacttattcc atgcattgtt ggcgttgcgtt tccttata
 15841 ttacggccgtt aagatataacc ataccccttta ctttgcgtt ggcgttgcgtt tccttata
 15901 ggggttctac atgcgttgcattt ctttgcgtt ggcgttgcgtt ggcgttgcgtt tccttata
 15961 tcgcacacgg cgcacgtt ggcgttgcgtt ggcgttgcgtt ggcgttgcgtt tccttata
 16021 cgagctgtatc cacagcttgc aaagggccctt ggcgttgcgtt ggcgttgcgtt tccttata
 16081 cgagcttgcattt tttgcgttgcattt ggcgttgcgtt ggcgttgcgtt tccttata
 16141 ggcgttgcgtt ggcgttgcgtt ggcgttgcgtt ggcgttgcgtt tccttata
 16201 cttggaggaa tatgacgggg acgatggat cttggaggaa gacggcgtt ggcgttgcgtt
 16261 gatgtttctt atccatgtatc gcaagacgc acggacccgg cttggaggaa gacggcgtt
 16321 agccagccgtt cccggccattttttt ctttgcgtt ggcgttgcgtt tccttata
 16381 tcgcgttgcattt ctttgcgtt ggcgttgcgtt tccttata
 16441 gcaattctgg aagcggttgcgtt cccggccgtt ggcgttgcgtt tccttata
 16501 atcgtaaaacggc cttggccgtt ggcgttgcgtt tccttata

17/92

16561 gacgcgctgc ttcaagcgcgt ggctcggtac aacagcagca acgtgcagac caacactggac
 16621 cggctgggtgg gggatgtgcg cgaggccgtg gcgcagcgtg agcgcgcga gcagcagggc
 16681 aacctgggtct ccatgggtgc actaaacgcgc ttcctgagta cacagccgc caacgtgccg
 16741 cggggacagg aggactacac caactttgtg agcgcactgc ggctaattgt gactgagaca
 16801 cgcggaaatgt aggtgtatca gtccgggcca gactatttt tccagaccag tagacaaggc
 16861 ctgcagacccg taaacctgag ccaggcttc aagaacttgc aggggctgtg ggggtgcgg
 16921 gctcccacag gcgaccgcgc gaccgtgtct agcttgctga cggccaaactc ggcctgttg
 16981 ctgctgctaa tagcgcctt cacggacagt ggcagcgtgt cccggacac ataccttaggt
 17041 cacttgcgtca cactgttaccg cgaggccata ggtcaggcgc atgtggacga gcatacttc
 17101 caggagatta caagtgttag cgcgcgtg gggcaggagg acacggcag cctggaggca
 17161 accctgtactt acctgcgtac caaccggcg caaaaaatcc cctcggtgca cagttaaac
 17221 agcgaggagg agcgcatttt ggcgtatgtc cagcagagcg tgagccttaa cctgatgcgc
 17281 gacggggtaa cggccagcgt ggcgtggac atgaccgcgc gcaacatgga accgggatg
 17341 tatgcctcaa accggccgtt tatcaatcgc ctaatggact acttgcacatcg cgcggccgc
 17401 gtgaaccccg agtattttcac caatgcacat ttaaaccgcg actggctacc gccccctgg
 17461 ttctacaccc ggggattcga ggtgcccgg ggttaacgtg gatttctgt ggacgacata
 17521 gacgcacgcg tggtttcccc gcaaccgcag accctgttag agttgcaaca acgcgagcag
 17581 gcagaggccg cgctgcgaaa ggaaagcttc cgcaggccaa gcagctgtc cgatctaggc
 17641 gctgcggccc cgcgggtcaga tgcttagtgc ccatttccaa gcttgatagg gtctttacc
 17701 agcactgcac ccacccggcc ggcgcgtgt ggcgaggagg agtacctaaa caactgcgt
 17761 ctgcagccgc agcgcgaaaaa gaacctgcct cgcgttgc ccaacaacgg gatacggagc
 17821 ctatggaca agatgagtag atggaagacg tatgcgcagg agcagacggg tggtccggc
 17881 cgcgcggccgc ccacccgtcg tcaaaaggcac gaccgtcagc ggggtctgt gtggaggac
 17941 gatgactcg gacacgcac cagcgtttt gatttggag ggagtggcaa cccgttgca
 18001 caccttcgccc ccaggctggg gagaatgtg taaaaaaaaa catgatgca aataaaaaaa
 18061 tcaccaaggc catggcaccg aegttgtt ttcttgtatt ccccttagta tgcggcgc
 18121 ggcgatgtat gagaaagggtc ctcccccctc ctacgagacg gtggtgagcg cggcccaagt
 18181 ggcggccgcg ctgggttac ctttcgtatgc tcccctggac ccccggtcg tgcctccgc
 18241 gtacctgcgg cttaccgggg ggagaaacacg catccgttac tctgagttgg caccctatt
 18301 cgacaccacc cgtgtgtacc ttgtggacaa caagtcaacg gatgtggcat ccctgaacta
 18361 ccagaacgcg cacagcaact ttctaaaccac ggtcattcaa aacaatgact acagccggg
 18421 ggaggcaagc acacagacca tcaatcttgc cgaccgtcg cactggggcg ggcacactgaa
 18481 aaccatccctg cataccaaca tgccaaatgt gaacgagttc atgtttacca ataagttaa
 18541 ggcgggggtg atgggtgtcgc gtcgttac taaggacaaa caggtggagc taaaatacga
 18601 gtgggtggat ttacgcgtc cccggggca ctactccgg accatgacca tagaccttat
 18661 gaacaacgcg atcgtggagc actacttgc agtggggcagg cagaacgggg ttctggaaag
 18721 cgacatcggtt gtaaaagtttgc acacccgcgg cttcagactg gggtttgacc cagtcactgg
 18781 tcttgtcatg cttggggtat atacaaacga agccttccat ccagacatca ttttgtgtcc
 18841 aggatgcggg gtggacttca cccacagccg cctgagcaac ttgttggca tccgcagcg
 18901 gcaacccttc caggagggtct ttaggatcac ctacgatgac ctggagggtg gtaacattcc
 18961 cgcactgttgc gatgtggacg cttaccaggc aagcttgc gatgacacccg aacagggccg
 19021 ggggtggcgca ggcggccgcgacaacatggc cagcggccgcg gaagagaact ccaacgcggc
 19081 agctgcggca atgcagccgg tggaggacat gaacgatcat gccattcgcc ggcacaccc
 19141 tgccacacgg gcccggggaga agcgcgtga ggcggaggca gcccggcaag ctgcccggcc
 19201 cgcactgttgc gtcgcacaac cccgggtcgaa gacccctcg aagaaacccg tgattaaacc
 19261 cctgacagag gacacgcaga aacgcagttt caacctaata agcaatgaca gcacccatc
 19321 ccagtaccgc agtgggtacc ttgcataaca ctacggccgc cctcaggccg ggatccgctc
 19381 atggaccctg ctttgcactc ctgacgtaa ctggggctcg gaggcaggat actggtcgtt
 19441 gcccgcacatg atgcacggacc cgcgttgc cccgtccacg cgcacatca gcaactttcc
 19501 ggtgggtggc gcccggatgt tgcccggtca ctccaaaggac ttctacaacg accaggccgt
 19561 ctactcccaag ctatccgcg agtttaccc tctgacccac gtgttcaatc gctttcccg
 19621 gaaccagatt ttggcgcgcg cgcggccccc caccatcacc accgtcagtg aaaacgttcc
 19681 tgctctcaca gatcacgggac cgcgttgc cccgtccacg cgcacatca gcaactttcc
 19741 gaccattact gacgcacccgac gcccgcacccccc cccctacgtt tacaaggccc tggcatagt
 19801 ctcgcgcgc gtcctatcga gcccgcacccccc ttgagcaacgc atgtccatcc ttatatcgcc

FIG. 4G

18/92

19861 cagcaataac acaggctggg gcctgcgtt cccaagcaag atgtttggcg gggccaagaa
 19921 ggcgtccgac caacacccag tgcgcgtcg cggcactac cgccgcgcct ggggcgcga
 19981 caaacgcggc cgcactggc gcaccacgt cgatgacgcc atcgacgcgg tggtgagga
 20041 ggcgcgcaac tacacgccc cgccgcgc agtgtccacc gtggacgcgg ccattcagac
 20101 cgtggtgcgc ggagcccgcc gctacgtta aatgaagaga cggcggaggg gcgttagcacf
 20161 tcgcccaccgc cgccgaccccg gcaactgcgc ccaacgcgcg gccggggccc tgcttaaccg
 20221 cgcacgtcgcc accggccgac gggccgcatt gcgagccgt cgaaggctgg cccgggttat
 20281 tgtcaactgtg ccccccaaggc ccaggcgacg agcggccgcg gcagcagcgc cggccattag
 20341 tgctatgact cagggtcgca ggggcaacgt gtactgggtg cgcgactcg tttagggcct
 20401 ggcgcgtccc gtgcgcaccc gccccccgcg caactagatt gcaataaaaa actacttaga
 20461 ctcgtactgt tgtatgtatc cagcggccgc ggcgcgcatac gaagctatgt ccaagcgaa
 20521 aatcaaagaa gagatgtcc aggtcatcgcc gccggagatc tatggccccc cgaagaagga
 20581 agagcaggat tacaagcccc gaaagctaaa gcccggcaaa aagaaaaaaga aagatgatga
 20641 tgcgtatgaa ctgcgcacg aggttggact gttgcacgcg accgcgcaca ggcacgggt
 20701 acagtggaaa ggtgcacgcg taagacgtt tttggcaccc ggcaccaccc tagtctttac
 20761 gcccggtgag cgctccaccc gcacctaaca ggcgtgtat gatggaggtt acggcgacga
 20821 ggacctgtt gaggcggcca acgagccct cggggaggtt gcctacggaa agcggcataa
 20881 ggacatgtg ggcgttgcgc tggacgaggg caacccaaca cctagcctaa agccctgtac
 20941 actgcagcag gtgcgtcccc cgcttgcacc gtccgaagaa aagcgcggcc taaagcgcga
 21001 gtctggtgcac ttggcacccca ccgtgcagct gatggtaccc aagcgtcagc gactggaaga
 21061 tgtcttgaa aaaatgaccg tggagccgtt gctggagccc gaggtccgcg tgccggcaat
 21121 caagcagggt gcacccggac tgggcgtca gaccgtggac gttcagatac ccaccacca
 21181 tagcactagt attgcactg ccacagaggg catggagaca caaacgtccc cggtgcctc
 21241 ggcgggtggca gatgcgcgcg tgcaggccgc cgctgcggcc gcttccaaga cctctacgga
 21301 ggtgcaaacg gaccgcgtt gtttgcgtt ttcagcccccc cggcgtccgc gccgtcaag
 21361 gaagtacggc gcccgcagcg cgctactgccc cgaatatgccc ctacatecctt ccatcgcgc
 21421 taccggccgc tatgcgtggc acacctaccc ccccaagaaga ctttttttttccggcc
 21481 aaccaccact ggaacccggc gcccgcgtcg ccgtcgcccg cccgtgttgg cccgatttc
 21541 cgtgcgcagg gtggcgtcg aaggaggcag gaccctgggtt ctgccaacag cgcgttacca
 21601 ccccgacatc gttaaaaaggc cggctttgtt ggttgcgtca gatatggccc tcacctgcgg
 21661 cctccgtttt ccgggtccgg gattccggg aagaatgcac cgtaggaggg gcatggccgg
 21721 ccacggcctg acggggcggca tgcgtcgatc gcaccacccgg cggcggcgcg cgtcgacc
 21781 tcgcgtcgcc ggcggtatcc tggccctctt tattccactg atgcggccgg cgattggcgc
 21841 cgtgcccggaa attgcacccg tggccttgc ggcgcagaga cactgattaa aaacaagtt
 21901 catgtggaaa aatcaaaaata aaagtctgg ctctcactg cgttgcgttcc tgtaactatt
 21961 ttgttagaatg gaagacatca actttgcgtc actggccccg cgacacggc cggccgggtt
 22021 catggggaaac tgcaagata tggcaccacg caatatgac ggtggcgcct tcagctgggg
 22081 ctcgcgttgg agcggcatta aaaatttccgg tcccgccgtt aagaactatg gcagcaaaaggc
 22141 ctggAACAGC agcacaggcc agatgcttag ggacaaggta aaagagcaaa atttccaaca
 22201 aaagggtggta gatggcctgg cctctggcat tagcgggggtt gtggacctgg ccaaccaggc
 22261 agtgcacaaat aagatataca gtaagcttgc tccccccctt cccgttagagg agcctccacc
 22321 ggcggggag acagtgtctc cagagggcgg tggcgaaaag cgtccgcgc cccgacaggga
 22381 agaaactctg gtacgcacaa tagacgcgc tccctcgatc gaggaggcgc taaagcaagg
 22441 cctggccacc acccgccatc tgcgcgcattt ggctaccggg gtgcgtggcc agcacacacc
 22501 cgtacgcgtg gacccgcctc ccccccgcga caccacggcgg aaacccgtgc tgccaggcccc
 22561 gtccggccgtt gtgttaaccc gtcctagccg cgcgtccctg cggccgcgc cccagggtcc
 22621 gcgatcggtt cggcccgtag ccagtggcaaa ctggcaaaagc acactgaaca gcatcggtgg
 22681 tttgggggtt caatccctga agcggccacg atgtttctga tagctaactgt gtcgtatgt
 22741 tgtcatgtat ggcgtccatgt cggccgcaga ggagctgtc agccgcgcg cggccgcctt
 22801 ccaagatggc taccggccatc atgatggccgc agtggctta catgcacatc tcggggccagg
 22861 acgcctcgaa gtacctgagc cccgggttgg tgcgttcgc cccgcgcacc gagacgtact
 22921 tcagcctgaa taacaagttt agaaacccca cggtggcgc tacgcacgc gtcgaccac
 22981 accggctctca gctggatc gtcgggttca tccccgttgg cccgcggat actgcgtact
 23041 cgtacaaggc gccggttcacc ctagctgtgg gtgataaccg tgcgtatgt gtcgaccac
 23101 cgtactttga catccgcggc gtgcgttgc gggccctac ttttaagccc tactctggca

FIG. 4H

19/92

23161 ctgcctacaa cgcactggcc cccaagggtg ccccaactc gtgcgagtgg gaacaaaatg
 23221 aaactgcaca agtggatgct caagaacttg acgaagagga gaatgaagcc aatgaagctc
 23281 aggcgcgaga acaggaacaa gctaagaaaa cccatgtata tgcccaggtt ccactgtccg
 23341 gaataaaaaat aactaaagaa ggtctacaaa taggaactgc cgacgccaca gttagcaggtg
 23401 cccgcaaaga aattttcgca gacaaaactt ttcaacactga accacaagta ggagaatctc
 23461 aatggAACGA agcggatgcc acagcagctg gtggagggt tcttaaaaag acaactccca
 23521 tggaaaccctg ctatggctca tacgctagac ccaccaattc caacggcgga cagggcgta
 23581 tgggtgaaca aaatggtaaa ttggaaagtc aagtgcgaaat gcaattttt tccacatcca
 23641 caaatgccac aaatgaagtt aacaatatac aaccaacagt tgtattgtac agcgaagatg
 23701 taaacatgga aactccagat actcatctt ctatataacc taaaatgggg gataaaaatg
 23761 ccaaagtcat gcttggacaa caagcaatgc caaacagacc aaattacatt gcttttagag
 23821 acaattttat tggtctcatg tattacaaca gcacaggtaa catgggtgtc cttgtggtc
 23881 aggcatcgca gttgaacgct gttgttagatt tgcacagacag aaacacagag ctgtccctacc
 23941 agcttttgc tgattcaatt ggcacagaaa caagataactt ttcaatgtgg aatcaagctg
 24001 ttgacagcta tgatccagat gtcagaatta ttgagaacca tggaaactgag gatgagttgc
 24061 caaattattt cttcccttctt ggtggatttgg ggtactgtc cacttttcaa gctgttaaaa
 24121 caactgctgc taacggggac caaggcaata ctacactggca aaaagattca acatttgcag
 24181 aacgcaatga aataggggtg gggaaataact ttgcacatggaa attaaacctg aatgccaacc
 24241 tatggagaaa ttccctttac tccaaatatttgc cgctgtaccc gccagacaag ctaaaataca
 24301 accccaccaa tggggaaata tctgacaacc ccaacacacta cgactacatg aacaagcgag
 24361 tgggtggtcc tgggttggta gactgtacaa ttaaccttgg ggcgcgtgg tctctggact
 24421 acatggacaaa ctttaaccacc accgcaatgc gggcctgcgt taccgctcca
 24481 tgggttggg aaacggccgc tacgtgcctt ttccacattca ggtggcccaa aagtttttgg
 24541 ccattaaaaa cttcccttctc ctggcaggct catacacata tgaatggaaac ttccagggaa
 24601 atgttaacat gtttctgcag agctcttgg gaaacgacact tagatgtgac ggggttagca
 24661 ttaaggttga cagcatttgt ctttacgcca ctttccccc catggcccac aacacggccct
 24721 ccacgcgtgg aaccatgctc agaaaatgaca ccaacgacca gtcctttaat gactacctt
 24781 ccggccccaa catgttatcat cccatacccg ccaacgccc caacgtgccc atctccatcc
 24841 catcgccaa ctgggcagca tttcgcggtt gggccttcac acgcttgcag acaaaggaaa
 24901 cccctccctt ggatcaggc tacgacccctt actacaccta ctctggctcc ataccatacc
 24961 ttgacggaaac ctcttatctt aatcacacactt ttaagaaggt ggccattact tttgactctt
 25021 ctgttagctg gcccggcaac gaccgcctgc ttactccca tgagttttagg attaagcgct
 25081 cagttgacgg ggagggtctat aacgtacgtc agtgcacat gacaaaggac tgggtccctag
 25141 tgcagatgtt ggcacactac aatattggctt accagggtttt ctacattcca gaaagctaca
 25201 aagacccatgttactcggttcc ttccagaaact tccagcccat gagccggcaa gtgggtggacg
 25261 atactaaata caaagattat cagcagggttgaattatcca ccagcataac aactcaggct
 25321 tcgttaggttcc cttcgctcccc accatgcgcg aggacaagc ttaccccttgc aatgtccct
 25381 acccactaat aggcaaaaacc gcgggtgata gtattacca gaaaaaggattt ctttgcgacc
 25441 gcacccctgtg ggcacatcccc ttctccagta actttatgtc catgggtgcgt ctcacagacc
 25501 tggggccaaaaa ctttctctac gcaaaactcccg cccacgcgt agacatgacc tttgagggtgg
 25561 atcccatggc cgagccacc ctttttatgt ttttgggttga agtctttgac tgggtccctg
 25621 tgcaccagcc gcacccgcgc gtcacgcgatccgcgttgcgttgcgcgc ttctcgcccg
 25681 gcaacgcac aacataaaaga agcaagcaac atcaacaaca gtcgcgcac tgggtccctc
 25741 tgagcaggaa ctgaaagccca ttgtcaaaaga ttttttttttttttttttttttttttttttttttt
 25801 ctatgacaag cgcttcccttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 25861 cacggccggc cgcgcgtactg ggggggttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 25921 aaaaacatgc taccttttttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 25981 ccagtttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 26041 tataacgcgtg gaaaaggccca cccaaaggcttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 26101 attctgtgc atgtttctcc acgccttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 26161 cccccccatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 26221 gcccacccctg cggcccaacc aggaacagcttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 26281 cttccgcggc cacagtgcgcg aaatttaggag cggccacttgcgttgcgttgcgttgcgttgcgtt
 26341 gtaaaaataaa tgacttagga gacacttca ataaaggcaaa atgtttttat ttgtacactc
 26401 tcgggtgatt atttacccccc acccttgcgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt

FIG. 41

20/92

26461 gccgcgcatac gctatgcgcc actggcaggg acacgttgcg atactggtgt ttagtgctcc
 26521 acttaaaactc aggcaacaacc atccgcgca gtcgggtgaa gtttcactc cacaggtgc
 26581 gcaccatcac caacgcgtt agcaggtcg ggcgcgatat cttgaagtgc cagttgggc
 26641 ctccgcctc cgccgcgcgag ttgcataca cagggttaca gcactggaaac actatcagcg
 26701 cgggtgggtg cacgcgtggcc agcacgtct tgcggagat cagatcccg tccaggtcct
 26761 cgcgttgct caggcgaaac ggagtcaact ttggtagctg cttcccgg aagggtgcac
 26821 gcccaggctt tgagttgcac tcgcaccgtt gtggcatcag aaggtgaccg tgcccagtct
 26881 gggcgtttagg atacagcgcc tgcataaag ccttgatctg cttaaaaggcc acctgagcct
 26941 ttgcgccttc agagaagaac atgcccgaag acttgcgggaa aactgtattt gccggacagg
 27001 ccgcgtcatg cacgcagcac ctgcgtcg tggtagat ctgcaccaca ttgcggcccc
 27061 accggttctt cagcatcttgc ctttgcgtt actgtccctt cagcgcgcgc tgcccgttt
 27121 cgctcgtcac atccattca atcactgtct ctttattttataatgtctc ccgttagac
 27181 acttaagctc gccttcgatc tcagcgcgc ggtgcagccaa acacgcgcag cccgtggct
 27241 cgtgggtctt ttaggttacc tctgcacaaacg actgcaggtt cgcctgcagg aatcgcccc
 27301 tcatcgtcac aaaggcttttgc tgcgtggta aggtcagctg caacccgcgg tgctccgt
 27361 ttagccaggcttgcatac gcccggcagg cttccacttgc tgcaggcagt agcttgaagt
 27421 ttgcctttag atcggttatcc acgtggtaact tgcctatcaa cgcgcgcgc gctccatgc
 27481 ctttctccca cgcagacacg atcggcaggc tcagcgggtt tattaccgtt ctttactt
 27541 cgcgttcaact ggacttccgc ttttccttgcatccgc acccgcggc actgggtcg
 27601 ctgcatttcgc cgcgcgcacc gtcgcgttac ctcccttgcgtt gtcgttgcgtt agcaccggc
 27661 ggttgcgaa acccaccatt tgcgcgcgc catcttccttgcgtt ctttccgtt ctgtccacga
 27721 tcacccctgg ggttgcgggg cgcgtgggtt tggaggggg ggcgttctt ttcttttgg
 27781 acgcaatggc caaatccgcgc tgcagggtcg atggccgcgg gtcgggtgtt cgcggccacca
 27841 ggcgcatttttgc tgcgcgttgc ttttcgttcc cggactcgatc acgcgcgcgc agccgtttt
 27901 ttgggggcgc gggggggaggc ggcggcgc ggcgcggggc gggacgttcc tccatgggg
 27961 gtggacgtcg cggccgcaccg cgtccggcgtt cgggggtgtt ttcgcgttc tcctttcc
 28021 gactggccat ttcccttc tataggcaga aaaagatcat ggagtcaatc gagaaggagg
 28081 acagccaaac cggcccccggc tgcgcgcgc cccggccctt caccatgcgc gccaacgcgc
 28141 ctaccacccctt cccgcgtcgag gcaccccccgc ttggaggagg ggaagtgtt atcgagcagg
 28201 acccagggtt tgcgcgcgc gacgcgcgcg atcgctcgat accaacagag gataaaaaagg
 28261 aagaccaggc cgcgcgcgc gcaaacgcgg aacaagtcgg gggggggggc caaaggcatg
 28321 ggcgcgttgc tgcgcgttgc gacgcgttc tgcgcgttgc tctgcgcgc cagtgcgc
 28381 ttatctgc
 28441 ttgccttccatgc
 28501 catgc
 28561 ccacccatca catcttttttgc ctttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
 28621 ggc
 28681 tcgcacgcgttgc
 28741 ctctgc
 28801 gtgcacgc
 28861 cggcacttgc
 28921 gtgcacgc
 28981 cggc
 29041 agcgcacgc
 29101 gtt
 29161 ggc
 29221 ctttccatgc
 29281 agggcgaggc ggc
 29341 ggcaacgc
 29401 agaagctgttgc
 29461 cgc
 29521 tgccagacttgc
 29581 cggaaatt
 29641 gtgaatgc
 29701 ctttccatgc

FIG. 4J

21/92

29761 gctgcaacct atgcaccccg caccgctccc tggctgcaa ttcacaactg cttagcgaaa
 29821 gtcaaattat cggtaccttt gagctgcagg gtcctcgcc tgacgaaaag tccgcggctc
 29881 cggggttgaa actcaactccg gggctgtgga cgtcggctt ctttcgaaaa ttgtacctg
 29941 aggactacca cggccacgag attaggttct acgaagacca atcccccccg ccaaattgcgg
 30001 agcttaccgc ctgcgtcatt acccagggcc acatccttgg ccaattgcaaa gccattaaca
 30061 aagccgcga agagttctg ctacgaaagg gacggggggg ttacttgac ccccaagtccg
 30121 gcgaggagct caacccaato ccccccgcgc cgcagcccta tcagcagccg cggggcccttg
 30181 cttcccaagga tggcacccaa aaagaagctg cagtcggccg cggccgcacc cacggacgag
 30241 gaggaaatct gggacagtca ggcagaggag gttttggac aggaggagga gatgatggaa
 30301 gactgggaca gcctagacga ggaagcttcc gaggccgaag aggtgtcaga cgaacacccg
 30361 tcaccctcgg tcgcatttcc ctcggccggc ccccaagaaat cggcaacccgt tccacgatt
 30421 gctacaaccc ctgcgtcctca ggccggccgg gcactgcccgg ttcggccgacc caaccgtaga
 30481 tgggacacca ctggaaaccag ggccggtaag tctaaggcgc cggccgggtt agccaaagag
 30541 caacaacacgc gccaaggcta ccgcctgtgg cgcgtcaca agaacgcctt agttgttgc
 30601 ttgcaagact gtggggggcaat ctctccctc gcccggccgt ttcttcata ccacacccgc
 30661 gtggcccttcc cccgttaat ctcgttac taccgtcatc tctacagccc ctactgcacc
 30721 ggcggcagcg gcagcaacag cagccggcac gcagaagcaa aggccggg atagaagac
 30781 tctgacaag cccaagaaat ccacagccgc ggcagcagca ggaggaggag cactgcgtct
 30841 ggcggccaaac gaaccctgtat cgcggccgaa gcttagaaac aggtttttc ccactctgt
 30901 tgctatattt caacagagca gggggcaaga acaagagctg aaaataaaaa acaggctct
 30961 ggcgtccctc acccgagct gcctgtatca caaaagcgaa gatcagctt ggcgcacgct
 31021 ggaagacgcg gaggctctt tcagaaata ctgcgcgtc actcttaagg actagttcg
 31081 cgcctttctt caaatthaag cgcggaaact acgtcatctc cagcggccac accggcgcc
 31141 agcacctgtc gtcagcgcca ttatgagcaa gggaaattccc acgccttaca tgtggagtt
 31201 ccagccacaa atgggacttg cggctggagc tgcccaagac tactcaaccc gaataaaacta
 31261 catgagcgcg ggaccccaaa tgatatcccc ggtcaacgaa atccgcgccc accggaaacccg
 31321 aattctccctc gaacaggccg ctattaccac cacacctgt aataaccta atccccgtag
 31381 ttggcccgct gcccctgggt accagggaaat tcccgctccc accactgtgg tacttcccag
 31441 agacgcccag gccgaagttc agatgactaa ctcaggggcg cagcttgcgg ggggtttcg
 31501 tcacaggggt cggtcgcccg ggcagggtat aactcacctg aaaatcagag ggcgaggat
 31561 tcagctcaac gacgagtccg tgagctcctc tcttggtctc cgtccggacg ggacattca
 31621 gatcggccgc gctggccgct cttcattac gcccgttag gcgatcctaa ctctgcagac
 31681 ctcgtccctc gagccgcgtc ccggaggcat tggactcta caatttattt aggagttcg
 31741 gccttcgggt tacttcaacc cttttctgg acctccggc cactacccgg accagtttat
 31801 tcccaacttt gacgcccgtaa aagactccgc ggacggctac gactgaatga ccagtggaga
 31861 ggcagagcaa ctgcgcctga cacacctcgaa ccactgcgcg cggccacaatgttggcccg
 31921 cggctccgggt gagttttgtt actttaattt gcccgaagag catatcgagg gcccggcgca
 31981 cggcgtccgg ctcaccaccc aggttagagct tacacgtac ctgattcggg agtttaccaa
 32041 ggcggccctg ctatgtggac gggaggggg tccctgtgtt ctgaccgtgg ttgtcaactg
 32101 tcctaaacctt ggattacatc aagatctt tccattcaac taacaataaa cacacaataa
 32161 attacttact taaaatctgt cagcaaatct ttgtccagct tattcagcat cacctccctt
 32221 ccctccctcc aactctggta ttccagcagc cttttctgt cgaactttctt ccaaagtctt
 32281 aatgggatgt caaattccctc atgttcttgt ccctccgcac ccactatctt catattgtt
 32341 cagatgaaac ggcgcagacc gtctgaagac accttcaacc ctgtgtaccc atatgacac
 32401 gaaacccggcc ctccaactgt gcctttccctt accccctccct ttgtgtcgcc aaatgggttc
 32461 caagaaagtc ccccccggagt gctttcttg cgtctttcag aacctttgtt tacctcac
 32521 ggcgtcttgc cgtaaaaaat gggcagccgc ctgtccctgg atcaggccagg caacccttaca
 32581 tcaaatacaa tcaactgtttc tcaaccgtt aaaaaaaaaa agtccaatata aactttggaa
 32641 acatccgcgc cccttacagt cagtcagggc gcccataacca tggccacaac ttgccttt
 32701 gtggctctg acaacactct taccatgca tcacaagcac cgctaaacctg gcaagactca
 32761 aaacttagca ttgttaccaa agagccactt acgtgttag atggaaaact ggccctgcag
 32821 acatcagccc ccctctctgc cactgataac aacgccttca ctatcactgc ctcacccct
 32881 cttactactg caaatggtag tctggctgtt accatggaaa accccacttta caacaacaat
 32941 gggaaacttg ggctaaaaat tggcggtcct ttgcaagtg ccaccgactc acatgcacta
 33001 acacttagtta ctgggtcaggg ggttgcagtt cataacaatt tgctacatc aaaagttaca

FIG. 4K

22/92

33061 ggccgcaatacg ggtttgatac atctggcaac atggaaactt aaactggaga tggcctctat
33121 gtggatagcg ccggctctaa ccaaaaacta catattaatc taaataccac aaaaggccct
33181 gctttgaca acaccgcaat aacaattaac gctggaaaag ggttgaatt taaaacagac
33241 tcctcaaagc gaaatccccat aaaaacaaaa attggatcatgc gcataacaata taataccaat
33301 ggagctatgg ttgcaaaact tggAACAGGC CTCAGTTG AGAGTCCGG AGCCATAACA
33361 atgggcagca taaacaatga cagacttaat cttggacaa caccagaccc atccccaaat
33421 tgcagaattt cttcagataa agactgcaag ctaactctgg cgctaacaaa atgtggcagt
33481 caaattttgg gcactgtttc agcttggca gtatcaggtt atatggcctc catcaatgg
33541 actctaagca gtgtaaaactt gggtcttaga tttgatgaca acggagtgt tatgtcaaat
33601 tcatacactgg acaaacagta ttggacttt agaaacgggg actccactaa cggtaaccca
33661 tacactttag ctgttgggtt tatgccaaac ctaaaagctt accccaaaaac tcaaagtaaa
33721 actgcaaaaaa gtaatattgt tagccaggtg tatcttaatg gtgacaagtc taaaccattt
33781 cattttacta ttacgctaaa tggaacagat gaaaccacc aagtaagcaa atactcaata
33841 tcattcagtt ggtcctggaa cagtggacaa tacactaatg acaaatttgc caccattcc
33901 tataccttct cctacattgc ccaggaataaa agaatcgtga acctgttgc tggttatgtt
33961 caacgtgtt attttcaat tgcaaaaaat ttcaagtcat ttttattca gtagtata
34021 cccaccacca catagcttactaatacacc gtaccttaat caaactcaca gaacccttagt
34081 attcaacctg ccacccctt cccaaacacac agagtagaca gtccttctc cccggctggc
34141 cttaaacagc atcatatcat gggtaacaga catattctt ggtttatat tccacacgg
34201 ctccgtcga gccaaacgct catcagtgat gttaaaaac tccccggca gtcgcttaa
34261 gttcatgtcg ctgtccagct gctgagccac aggctgtgt ccaacttgcg gttgtcaac
34321 gggcggcga ggagaagtcc acgcctacat ggggttagag tcataatcgt gcatcaggat
34381 agggcggtg tgctcagca ggcgcgcaat aaactgctgc cgccggcgcg cctgttgc
34441 ggaatacacaac atggcagtgg tcttcctcagc gatgattcgc accggccgca gcataaggcg
34501 ctttgtcctc cgggcacagc aggcacccct gatctcaactt aagttagcactc actaactgca
34561 gcacagtacc acaatattgt taaaatccc acagtgcgag ggcgtgttacttcaatgc
34621 ggcggggacc acagaaccca cgtggccatc ataccacaag cgcaggtaga ttaagtggcg
34681 acccctcata aacacgttgg acataaaatc taccttctt ggcatgttgc aattcaccac
34741 ctccggta catataaacc tctgattaaa catggcgc tccaccacca ctctttaacc
34801 gctggccaaa acctgcccgc cggctatgc ctcaggaa ccggacttgg acaatgaca
34861 gtggagagcc caggactcgt aaccatggat catatgcgc gtcatgatcatatgttgg
34921 acaacacagg cacacgtca tacacttctt caggattaca agtcctccc ggcgcgaaac
34981 catatccccag ggaacaaaccc atttctgaat cagcttaat cccacactgc agggaaagacc
35041 tcgcacgtaa ctcacgttgc gcattgtcaa agtggatcat tcggcagca gggatgatc
35101 ctccagttatg gtgcgcggg tttctgtctc aaaaggaggt agacatccc tactgtacgg
35161 agtgcgcgg gacaacccgg atcgtgttgg tcgttagtgc atgcataatg gaaacgcgg
35221 cgtatcata tttctgttgg caaaaccagg tgccggcgtg acaaacagat ctgcgtctcc
35281 ggtctcgccg cttagatcgc tctgtgttagt aactccctc atgcgcgcg gcccgtata
35341 ccaggcgcccc cctggcttcg gtttctatgt aaactccctc acatcgatc tgcgactc
35401 catccaccac cgcagaataa gccacacccca gccaacccatc acatcgatc tgcgactc
35461 acacggggagg agcggggaaa gctggaaagaa ccatgttttt ttttttattc caaaagatta
35521 tccaaaaccc taaaatgaag atcttataag tgaacgcgt cccctccgg ggcgtggta
35581 aactctacag caaaagaaca gataatggca ttgtaaatg gttgcacaat ggcttccaaa
35641 aggcaaacgg ccctcactcgc caagtggacg taaaggctt accccctcagg gtgaatctcc
35701 tctataaaaca ttccagcacc ttcaaccatg cccaaataat tctcatctcg ccacccctc
35761 aatatatctc taagcaatac cggaaatattt agtccggca ttgtaaaaat ctgtccaga
35821 ggcgcctcca cttcagctt ccaagcgcgca atcatgattt caaaattca gtttctc
35881 agacctgtat aagattcaaa agcggaaacat taacaaaaat acccgatcc cgttagtccc
35941 ttgcggggc cagctgaaca taatcgtca ggtctgcacg gaccagcgcg gccacttccc
36001 cgccaggaac catgcaaaaaa gaacccacac tgattatgc acgcataactc ggagctatgc
36061 taaccagcgt agccccatg taagttgtt gcatggccg cgatataaaa tgcaagggt
36121 tgctaaaaaa atcaggcaaa gcctcgcga aaaaagaaaac cacatcgat tcatgtcat
36181 gcagataaaag gcaggtaaagc tccggaaacca ccacagaaaa agacaccatt tttctctcaa
36241 acatgtctgc gggttctgc ataaaacacaa aataaaataa caaaaaaaaa tttaaacatt
36301 aqaaggctgt cttacaacag gaaaaacacac ctttataagc ataagacgggatc acggccat

FIG. 4L

23/92

36361 gcccggcgtga ccgtaaaaaa actggtcacc gtgattaaaa agcaccaccc acagctcctc
36421 ggtcatgtcc ggagtctaaa tgtaagactc ggtaaaacaca tcagggttat tcacatcggt
36481 cagtgtctaa aagcgaccga aatagccccgg gggaaatacat acccgccaggc gtagagacaa
36541 cattacagcc cccataggag gtataacaaa attaatagga gagaaaaaaca cataaacacc
36601 tgaaaaaacc tccctgcctag gcaaaaatagc accctccgc tccagaacaa catacagcgc
36661 ttcccacagcg gcagccataa cagtcaggct taccagtaaa aaagaaaaacc tattaaaaaa
36721 acaccactcg acacggcacc agctcaatca gtcacagtgt aaaaaaggc caagtgcaga
36781 gcgagttat ataggactaa aaaatgacgt aacggttaaa gtccacaaaa aacacccaga
36841 aaacccgcacg cgaacctacg cccagaaacg aaagccaaaa aacccacaaac ttccctcaaatt
36901 cgtaacttcc gttttccac gttacgtcac ttccatctt aagaaaaacta caattcccaa
36961 cacatacaag ttactccgcc ctaaaaaccta cgtcaccgc cccgttccca cgccccgcgc
37021 cacgtcacaac actccacccc ctcattatca tattggcttc aatccaaaat aaggtatatt
37081 attgatgatg

FIG. 4M

24/92

10 ATGGCGCCCATCACGGCCTACTCCAACAGACGCGGGGCCTACTTGGTTGCATCATCACT -----+-----+-----+-----+-----+ Met Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr	30 -----+-----+-----+-----+ 10 20	50
70 AGCCTTACAGGCCGGGACAAGAACGAGTCGAGGGAGAGGTTCAGGTGGTTCCACCGCA -----+-----+-----+-----+-----+ Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala	90 -----+-----+-----+-----+ 30 40	110
130 ACACAATCCTCCTGGCGACCTGCGTCAACGGCGTGTGTTGGACC GTT ACC ATGGTGCT -----+-----+-----+-----+-----+ Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala	150 -----+-----+-----+-----+ 50 60	170
190 GGCTCAAAGACCTTAGCCGGCCAAAGGGGCCAATCACCCAGATGTACACTAATGTGGAC -----+-----+-----+-----+-----+ Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp	210 -----+-----+-----+-----+ 70 80	230
250 CAGGACCTCGTCGGCTGGCAGGCGCCCCCGGGCGC GTT CTT GAC ACC ATGC ACCT GT -----+-----+-----+-----+-----+ Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys	270 -----+-----+-----+-----+ 90 100	290
310 GGCAGCTCAGACCTTACTTGGTCACGAGACATGCTGACGTCATTCCGGTGCGCCGGCG -----+-----+-----+-----+-----+ Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg	330 -----+-----+-----+-----+ 110 120	350
370 GGCGACAGTAGGGGGAGCCTGCTCTCCCCCAGGCCCTGTCCTACTTGAAGGGCTCTCG -----+-----+-----+-----+-----+ Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser	390 -----+-----+-----+-----+ 130 140	410

FIG. 5A

25/92

430	450	470
GGTGGTCCACTGCTCTGCCCTCGGGGCACGCTGTGGGCATCTCCGGGCTGCCGTATGC		
-----+-----+-----+-----+-----+-----+		
GlyGlyProLeuLeuCysProSerGlyHisAlaValGlyIlePheArgAlaAlaValCys		
150		160
490	510	530
ACCCGGGGGGTTGCGAAGGCGGTGGACTTGTGCCCGTAGAGTCATGGAAACTACTATG		
-----+-----+-----+-----+-----+-----+		
ThrArgGlyValAlaLysAlaValAspPheValProValGluSerMetGluThrThrMet		
170		180
550	570	590
CGGTCTCCGGTCTTCACGGACAACTCATCCCCCCC GGCGTACCGCAGTCATTCAAGTG		
-----+-----+-----+-----+-----+-----+		
ArgSerProValPheThrAspAsnSerSerProProAlaValProGlnSerPheGlnVal		
190		200
610	630	650
GCCCACCTACACGCTCCACTGGCAGCGGCAAGAGTACTAAAGTGCCGGCTGCATATGCA		
-----+-----+-----+-----+-----+-----+		
AlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAla		
210		220
670	690	710
GCCCAAGGGTACAAGGTGCTCGTCTCAATCCGTCCGTTGCCGCTACCTTAGGGTTGGG		
-----+-----+-----+-----+-----+-----+		
AlaGlnGlyTyrLysValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGly		
230		240
730	750	770
GCGTATATGTCTAACGGCACACGGTATTGACCCCAACATCAGAACTGGGTAAGGACCATT		
-----+-----+-----+-----+-----+-----+		
AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle		
250		260
790	810	830
ACCACAGGC GCCCCCGTCACATACTCTACCTATGGCAAGTTCTTGCCGATGGTGGTTGC		
-----+-----+-----+-----+-----+-----+		
ThrThrGlyAlaProValThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys		
270		280

FIG. 5B

26/92

850	870	890
TCTGGGGCGCTTATGACATCATAATATGTGATGAGTGCCATTCAACTGACTCGACTACA		
-----+-----+-----+-----+-----+-----+		
SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspSerThrThr		
290		300
910	930	950
ATCTTGGGCATCGGCACAGTCCTGGACCAAGCGGAGACGGCTGGAGCGCGGCTTGTCTG		
-----+-----+-----+-----+-----+-----+		
IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal		
310		320
970	990	1010
CTCGCCACCGCTACGCCCTCCGGATCGGTACCGTGCCACACCCAAACATCGAGGAGGTG		
-----+-----+-----+-----+-----+-----+		
LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal		
330		340
1030	1050	1070
GCCCTGTCTAATACTGGAGAGATCCCCTCTATGGCAAAGCCATCCCCATTGAAGCCATC		
-----+-----+-----+-----+-----+-----+		
AlaLeuSerAsnThrGlyGluIleProPheTyrGlyLysAlaIleProIleGluAlaIle		
350		360
1090	1110	1130
AGGGGGGGAAGGCATCTCATTTCTGTCATTCCAAGAAGAAGTGCACGAGCTGCCGCA		
-----+-----+-----+-----+-----+-----+		
ArgGlyGlyArgHisLeuIlePheCysHisSerLysLysCysAspGluLeuAlaAla		
370		380
1150	1170	1190
AAGCTGTCAGGCCTCGGAATCAACGCTGTGGCGTATTACCGGGGCTCGATGTGTCCGTC		
-----+-----+-----+-----+-----+-----+		
LysLeuSerGlyLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal		
390		400
1210	1230	1250
ATACCAACTATCGGAGACGTCGTTGTCGTGGCACACAGACGCTCTGATGACGGGCTATACG		
-----+-----+-----+-----+-----+-----+		
IleProThrIleGlyAspValValValAlaThrAspAlaLeuMetThrGlyTyrThr		
410		420

FIG. 5C

27/92

1270	1290	1310
GGCGACTTGTACTCAGTGATCGACTGTAACACATGTGTCACCCAGACAGTCGACTTCAGC		
-----+-----+-----+-----+-----+-----+		
GlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrValAspPheSer		
430		440
1330	1350	1370
TTGGATCCCACCTTCACCATTGAGACGACGACCGTGCCCTCAAGACGCAGTGTGCGCTCG		
-----+-----+-----+-----+-----+-----+		
LeuAspProThrPheThrIleGluThrThrValProGlnAspAlaValSerArgSer		
450		460
1390	1410	1430
CAGCGGCCGGGTAGGACTGGCAGGGTAGGAGAGGCATCTACAGGTTGTGACTCCGGGA		
-----+-----+-----+-----+-----+-----+		
GlnArgArgGlyArgThrGlyArgGlyArgGlyIleTyrArgPheValThrProGly		
470		480
1450	1470	1490
GAACGGCCCTCGGGCATGTTGATTCCCTCGGTCTGTGAGTGCTATGACGCCGGCTGT		
-----+-----+-----+-----+-----+-----+		
GluArgProSerGlyMetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCys		
490		500
1510	1530	1550
GCTTGGTACGAGCTCACCCCCGCCGAGACCTCGGTTAGGTTGCGGGCTACCTGAACACA		
-----+-----+-----+-----+-----+-----+		
AlaTrpTyrGluLeuThrProAlaGluThrSerValArgLeuArgAlaTyrLeuAsnThr		
510		520
1570	1590	1610
CCAGGGTTGCCCGTTGCCAGGACCACCTGGAGTTCTGGGAGAGTGTCTTCACAGGCCTC		
-----+-----+-----+-----+-----+-----+		
ProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluSerValPheThrGlyLeu		
530		540
1630	1650	1670
ACCCACATAGATGCACACTTCTTGTCCCAGACCAAGCAGGCAGGAGACAACCTCCCTAC		
-----+-----+-----+-----+-----+-----+		
ThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnAlaGlyAspAsnPheProTyr		
550		560

FIG. 5D

28/92

1690	1710	1730
CTGGTAGCATACCAAGCCACGGTGTGCGCCAGGGCTCAGGCCACCTCCATCATGGAT		
-----+-----+-----+-----+-----+		
LeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProSerTrpAsp		
570		580
1750	1770	1790
CAAATGTGGAAGTGTCTCATACGGCTGAAACCTACGCTGCACGGGCCAACACCCCTTGCTG		
-----+-----+-----+-----+-----+		
GlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeu		
590		600
1810	1830	1850
TACAGGCTGGGAGCCGTCCAAAATGAGGTCACCCTCACCCACCCCATAACCAAATACATC		
-----+-----+-----+-----+-----+		
TyrArgLeuGlyAlaValGlnAsnGluValThrLeuThrHisProIleThrLysTyrIle		
610		620
1870	1890	1910
ATGGCATGCATGTCGGCTGACCTGGAGGTGTCACTAGCACCTGGGTGCTGGTGGCGGA		
-----+-----+-----+-----+-----+		
MetAlaCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGly		
630		640
1930	1950	1970
GTCCTTGCAGCTCTGGCCCGTATTGCCTGACAACAGGCAGTGTGGTCATTGTGGTAGG		
-----+-----+-----+-----+-----+		
ValLeuAlaAlaLeuAlaAlaTyrCysLeuThrThrGlySerValValIleValGlyArg		
650		660
1990	2010	2030
ATTATCTTGTCCGGGAGGCCGGCTATTGTTCCCCGACAGGGAGTTCTCTACCAAGGAGTTC		
-----+-----+-----+-----+-----+		
IleIleLeuSerGlyArgProAlaIleValProAspArgGluPheLeuTyrGlnGluPhe		
670		680
2050	2070	2090
GATGAAATGGAAGAGTGCGCCTCGCACCTCCCTACATCGAGCAGGGAATGCAGCTGCC		
-----+-----+-----+-----+-----+		
AspGluMetGluGluCysAlaSerHisLeuProTyrIleGluGlnGlyMetGlnLeuAla		
690		700

FIG. 5E

29/92

2110	2130	2150
GAGCAATTCAAGCAGAAAGCGCTGGGTTACTGCAAACAGCCACCAACAAGCGGAGGCT -----+-----+-----+-----+-----+ GluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaThrLysGlnAlaGluAla 		
	710	720
2170	2190	2210
GCTGCTCCCCTGGTGGAGTCCAAGTGGCGAGCCCTTGAGACATTCTGGCGAAGCACATG -----+-----+-----+-----+-----+ AlaAlaProValValGluSerLysTrpArgAlaLeuGluThrPheTrpAlaLysHisMet 		
	730	740
2230	2250	2270
TCCAATTTCATCAGCGGGATAACAGTACTTAGCAGGCTTATCCACTCTGCCTGGGAACCCC -----+-----+-----+-----+-----+ TrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnPro 		
	750	760
2290	2310	2330
GCAATAGCATCATTGATGGCATTACAGCCTCTATCACCAAGCCCCGCTCACCAACCAAAGT -----+-----+-----+-----+-----+ AlaIleAlaSerLeuMetAlaPheThrAlaSerIleThrSerProLeuThrThrGlnSer 		
	770	780
2350	2370	2390
ACCCTCCTGTTAACATCTGGGGGGGTGGCTGCCAACTCGCCCCCCCCCAGCGCC -----+-----+-----+-----+-----+ ThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeuAlaProProSerAla 		
	790	800
2410	2430	2450
GCTTCGGCTTTCGTGGCGCCGGCATCGCCGGTGCAGCAGCATAGGCCTTGGG -----+-----+-----+-----+-----+ AlaSerAlaPheValGlyAlaGlyIleAlaGlyAlaAlaValGlySerIleGlyLeuGly 		
	810	820
2470	2490	2510
AAGGTGCTTGTGGACATTCTGGCGGGTTATGGAGCAGGAGTGGCCGGCGCTCGTGGCC -----+-----+-----+-----+-----+ LysValLeuValAspIleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAla 		
	830	840

FIG. 5F

30/92

2530	2550	2570
TTCAAGGTATGAGCGGGAGATGCCCTCCACCGAGGACCTGGTCAATCTACTTCCTGCC		
-----+-----+-----+-----+-----+		
PheLysValMetSerGlyGluMetProSerThrGluAspLeuValAsnLeuLeuProAla		
850		860
2590	2610	2630
ATCCTCTCTCCTGGCGCCCTGGTCGTCGGGTCGTGTGCAGCAATACTGCGTCGACAC		
-----+-----+-----+-----+-----+		
IleLeuSerProGlyAlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHis		
870		880
2650	2670	2690
GTGGGTCCGGGAGAGGGGGCTGTGCAGTGGATGAACCGGCTGATAGCGTTGCCCTCGCGG		
-----+-----+-----+-----+-----+		
ValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArg		
890		900
2710	2730	2750
GGTAATCATGTTCCCCACGCACATGTGCCTGAGAGCGACGCCGCAGCGCGTGTACT		
-----+-----+-----+-----+-----+		
GlyAsnHisValSerProThrHisTyrValProGluSerAspAlaAlaAlaArgValThr		
910		920
2770	2790	2810
CAGATCCTCTCCAGCCTTACCATCACTCAGCTGCTGAAAGGCTCCACCAGTGGATTAAAT		
-----+-----+-----+-----+-----+		
GlnIleLeuSerSerLeuThrIleThrGlnLeuLeuLysArgLeuHisGlnTrpIleAsn		
930		940
2830	2850	2870
GAAGACTGCTCACACCGTGTCCGGCTCGTGGCTAAGGGATGTTGGACTGGATATGC		
-----+-----+-----+-----+-----+		
GluAspCysSerThrProCysSerGlySerTrpLeuArgAspValTrpAspTrpIleCys		
950		960
2890	2910	2930
ACGGTGTTGACTGACTCAAGACCTGGCTCCAGTCCAAGCTCCTGCCAGCTACCGGGA		
-----+-----+-----+-----+-----+		
ThrValLeuThrAspPheLysThrTrpLeuGlnSerLysLeuLeuProGlnLeuProGly		
970		980

FIG. 5G

31/92

2950	2970	2990
GTCCCTTTTCTCGCAACGCGGGTACAAGGGAGTCTGGCGGGAGACGGCATCATG -----+-----+-----+-----+-----+-----+		
ValProPhePheSerCysGlnArgGlyTyrLysGlyValTrpArgGlyAspGlyIleMet 990 1000		
3010	3030	3050
CAAACCACCTGCCATGTGGAGCACAGATCACCGACATGTCAAAAACGGTTCCATGAGG -----+-----+-----+-----+-----+-----+		
GlnThrThrCysProCysGlyAlaGlnIleThrGlyHisValLysAsnGlySerMetArg 1010 1020		
3070	3090	3110
ATCGTCGGGCCTAACGACCTGCAGCAACACGTGGCATGGAACATTCCCCATCAACGCATAC -----+-----+-----+-----+-----+-----+		
IleValGlyProLysThrCysSerAsnThrTrpHisGlyThrPheProIleAsnAlaTyr 1030 1040		
3130	3150	3170
ACCACGGGCCCCCTGCACACCCCTCTCCAGCGCCAAACTATTCTAGGGCGCTGTGGCGGGTG -----+-----+-----+-----+-----+-----+		
ThrThrGlyProCysThrProSerProAlaProAsnTyrSerArgAlaLeuTrpArgVal 1050 1060		
3190	3210	3230
GCCGCTGAGGAGTACGTGGAGGTACCGCAGGTTCCGGCTCTGAATTCTCACGGACGGCATG -----+-----+-----+-----+-----+-----+		
AlaAlaGluGluTyrValGluValThrArgValGlyAspPheHisTyrValThrGlyMet 1070 1080		
3250	3270	3290
ACCACTGACAACGTAAAGTGCCCCATGCCAGGTTCCGGCTCTGAATTCTCACGGAGGTG -----+-----+-----+-----+-----+-----+		
ThrThrAspAsnValLysCysProCysGlnValProAlaProGluPhePheThrGluVal 1090 1100		
3310	3330	3350
GACGGAGTGGGTTGCACAGGTACGCTCCGGTGCAGGCCTCTCCTACGGGAGGAGTT -----+-----+-----+-----+-----+-----+		
AspGlyValArgLeuHisArgTyrAlaProAlaCysArgProLeuLeuArgGluGluVal 1110 1120		

FIG. 5H

32/92

3370	3390	3410
ACATTCCAGGTGGGCTCAACCAATACTGGTTGGTCACAGCTACCATGCGAGCCGAA -----+-----+-----+-----+-----+-----+		
ThrPheGlnValGlyLeuAsnGlnTyrLeuValGlySerGlnLeuProCysGluProGlu 1130 1140		
3430	3450	3470
CCGGATGTAGCAGTGCTCACTTCCATGCTCACCGACCCCTCCCACATCACAGCAGAACG -----+-----+-----+-----+-----+-----+		
ProAspValAlaValLeuThrSerMetLeuThrAspProSerHisIleThrAlaGluThr 1150 1160		
3490	3510	3530
GCTAAGCGTAGGTTGGCCAGGGGGTCTCCCCCTCCTGGCCAGCTCTCAGCTAGCCAG -----+-----+-----+-----+-----+-----+		
AlaLysArgArgLeuAlaArgGlySerProProSerLeuAlaSerSerSerAlaSerGln 1170 1180		
3550	3570	3590
TTGTCTGCGCCTTCCTGAAGGCGACATGCACTACCCACCATGTCTCTCCGGACGCTGAC -----+-----+-----+-----+-----+-----+		
LeuSerAlaProSerLeuLysAlaThrCysThrThrHisHisValSerProAspAlaAsp 1190 1200		
3610	3630	3650
CTCATCGAGGCCAACCTCCTGTGGCGGAGATGGCGGGAACATCACCGCGTGGAG -----+-----+-----+-----+-----+-----+		
LeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGlu 1210 1220		
3670	3690	3710
TCGGAGAACAGGTGGTAGTCCTGGACTCTTCGACCCGCTCGAGCGGAGGAGATGAG -----+-----+-----+-----+-----+-----+		
SerGluAsnLysValValValLeuAspSerPheAspProLeuArgAlaGluGluAspGlu 1230 1240		
3730	3750	3770
AGGGAAAGTATCCGTTCCGGCGGAGATCCTGCAGAAATCCAAGAAGTTCCCCGCAGCGATG -----+-----+-----+-----+-----+-----+		
ArgGluValSerValProAlaGluIleLeuArgLysSerLysLysPheProAlaAlaMet 1250 1260		

FIG. 51

33/92

3790	3810	3830
CCCATCTGGGCGCGCCCCGATTACAACCCCTCCACTGTTAGAGTCCTGGAAAGGACCCGGAC		
-----+-----+-----+-----+-----+-----+		
ProIleTrpAlaArgProAspTyrAsnProProLeuLeuGluSerTrpLysAspProAsp		
1270		1280
3850	3870	3890
TACGTCCCTCCGGTGGTGCACGGGTGCCCGTTGCCACCTATCAAGGCCCCCTCCAATACCA		
-----+-----+-----+-----+-----+-----+		
TyrValProProValValHisGlyCysProLeuProProIleLysAlaProProIlePro		
1290		1300
3910	3930	3950
CCTCCACGGAGAAAGAGGACGGTTGCCTAACAGAGTCCTCCGTGTCTCTGCCTTAGCG		
-----+-----+-----+-----+-----+-----+		
ProProArgArgLysArgThrValValLeuThrGluSerSerValSerSerAlaLeuAla		
1310		1320
3970	3990	4010
GAGCTCGCTACTAACGACCTTCGGCAGCTCCGAATCATGGCCGTCGACAGCGGCACGGCG		
-----+-----+-----+-----+-----+-----+		
GluLeuAlaThrLysThrPheGlySerSerGluSerSerAlaValAspSerGlyThrAla		
1330		1340
4030	4050	4070
ACCGCCCTCCTGACCAGGCCTCCGACGACGGTGACAAAGGATCCGACGGTTGAGTCGTAC		
-----+-----+-----+-----+-----+-----+		
ThrAlaLeuProAspGlnAlaSerAspAspGlyAspLysGlySerAspValGluSerTyr		
1350		1360
4090	4110	4130
TCCTCCATGCCCGCCCTTGAGGGGAACCGGGGGACCCCGATCTCAGTGACGGGTCTGG		
-----+-----+-----+-----+-----+-----+		
SerSerMetProProLeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrp		
1370		1380
4150	4170	4190
TCTACCGTGAGCGAGGAAGCTAGTGAGGATGTCGTCTGCTCAATGTCTACACATGG		
-----+-----+-----+-----+-----+-----+		
SerThrValSerGluGluAlaSerGluAspValValCysCysSerMetSerTyrThrTrp		
1390		1400

FIG. 5J

34/92

4210	4230	4250
ACAGGGCGCTTGATCACGCCATGCGCTGCGGAGGAAAGCAAGCTGCCATCAACGCGTTG -----+-----+-----+-----+-----+-----+		
ThrGlyAlaLeuIleThrProCysAlaAlaGluGluSerLysLeuProIleAsnAlaLeu 1410 1420		
4270	4290	4310
AGCAACTCTTGCTGCCACCATAACATGGTTATGCCACAACATCTCGCAGCGCAGGC -----+-----+-----+-----+-----+-----+		
SerAsnSerLeuLeuArgHisHisAsnMetValTyrAlaThrThrSerArgSerAlaGly 1430 1440		
4330	4350	4370
CTGCGGCAGAAGAACGGTCACCTTGACAGACTGCAAGTCCTGGACGACCACTACCGGGAC -----+-----+-----+-----+-----+-----+		
LeuArgGlnLysLysValThrPheAspArgLeuGlnValLeuAspAspHisTyrArgAsp 1450 1460		
4390	4410	4430
GTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTTAAGGCTAAACTCCTATCCGTAGAG -----+-----+-----+-----+-----+-----+		
ValLeuLysGluMetLysAlaLysAlaSerThrValLysAlaLysLeuLeuSerValGlu 1470 1480		
4450	4470	4490
GAAGCCTGCAAGCTGACGCCACATTGGCCAAATCCAAGTTGGCTATGGGGCAAAG -----+-----+-----+-----+-----+-----+		
GluAlaCysLysLeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAlaLys 1490 1500		
4510	4530	4550
GACGTCCGGAACCTATCCAGCAAGGCCGTTAACCATCCACTCCGTGTGGAAGGACTTG -----+-----+-----+-----+-----+-----+		
AspValArgAsnLeuSerSerLysAlaValAsnHisIleHisSerValTrpLysAspLeu 1510 1520		
4570	4590	4610
CTGGAAGACACTGTGACACCAATTGACACCACATGGAAAAAATGAGGTTCTGT -----+-----+-----+-----+-----+-----+		
LeuGluAspThrValThrProIleAspThrThrIleMetAlaLysAsnGluValPheCys 1530 1540		

FIG. 5K

35/92

4630	4650	4670
GTCCAACCAGAGAAAGGAGGCCGTAAGCCAGCCCGCCTTATCGTATTCCCAGATCTGGGA		
-----+-----+-----+-----+-----+-----+		
ValGlnProGluLysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeuGly		
1550 1560		
4690	4710	4730
GTCCGTGTATGCGAGAAGATGGCCCTCTATGATGTGGTCTCCACCCCTCAGGTCGTG		
-----+-----+-----+-----+-----+-----+		
ValArgValCysGluLysMetAlaLeuTyrAspValValSerThrLeuProGlnValVal		
1570 1580		
4750	4770	4790
ATGGGCTCCTCATACGGATTCCAGTACTCTCCTGGGCAGCGAGTCGAGTTCCCTGGTGAAT		
-----+-----+-----+-----+-----+-----+		
MetGlySerSerTyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuValAsn		
1590 1600		
4810	4830	4850
ACCTGGAAATCAAAGAAAAACCCCATGGGCTTTCATATGACACTCGCTGTTGACTCA		
-----+-----+-----+-----+-----+-----+		
ThrTrpLysSerLysLysAsnProMetGlyPheSerTyrAspThrArgCysPheAspSer		
1610 1620		
4870	4890	4910
ACGGTCACCGAGAACGACATCCGTGTTGAGGAGTCATTTACCAATGTTGACTTGGCC		
-----+-----+-----+-----+-----+-----+		
ThrValThrGluAsnAspIleArgValGluGluSerIleTyrGlnCysCysAspLeuAla		
1630 1640		
4930	4950	4970
CCCGAAGCCAGACAGGCCATAAAATCGCTCACAGAGCGGCTTATATCGGGGGTCCCTTG		
-----+-----+-----+-----+-----+-----+		
ProGluAlaArgGlnAlaIleLysSerLeuThrGluArgLeuTyrIleGlyGlyProLeu		
1650 1660		
4990	5010	5030
ACTAATTCAAAAGGGCAGAACTGCGGTTATCGCCGGTGCCCGCGAGCGGCGTGCTGACG		
-----+-----+-----+-----+-----+-----+		
ThrAsnSerLysGlyGlnAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeuThr		
1670 1680		

FIG. 5L

36/92

5050	5070	5090
ACTAGCTGCGGTAACACCCTCACATGTTACTTGAAGGCCTCTGCAGCCTGTCGAGCTGCG		
-----+-----+-----+-----+-----+		
ThrSerCysGlyAsnThrLeuThrCysTyrLeuLysAlaSerAlaAlaCysArgAlaAla		
1690 1700		
5110	5130	5150
AAGCTCCAGGACTGCACGATGCTCGTAACGGAGACGACCTTGTGTTATCTGTGAAAGC		
-----+-----+-----+-----+-----+		
LysLeuGlnAspCysThrMetLeuValAsnGlyAspAspLeuValValIleCysGluSer		
1710 1720		
5170	5190	5210
GCGGGAACCCAAGAGGACGCGGCCGAGCCTACGAGTCTTCACGGAGGCTATGACTAGGTAC		
-----+-----+-----+-----+-----+		
AlaGlyThrGlnGluAspAlaAlaSerLeuArgValPheThrGluAlaMetThrArgTyr		
1730 1740		
5230	5250	5270
TCTGCCCCCCCCGGGGACCCGCCAACAGAACATCGACTTGGAGCTGATAACATCATGT		
-----+-----+-----+-----+-----+		
SerAlaProProGlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSerCys		
1750 1760		
5290	5310	5330
TCCTCCAATGTGTCGGTCGCCACGATGCATCAGGCAAAAGGGTGTACTACCTCACCGT		
-----+-----+-----+-----+-----+		
SerSerAsnValSerValAlaHisAspAlaSerGlyLysArgValTyrTyrLeuThrArg		
1770 1780		
5350	5370	5390
GATCCCACCACCCCCTCGCACGGGCTGCGTGGGAAACAGCTAGACACACTCCAGTTAAC		
-----+-----+-----+-----+-----+		
AspProThrThrProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProValAsn		
1790 1800		
5410	5430	5450
TCCTGGCTAGGCAACATTATCATGTATGCGCCACTTGTGGCAAGGATGATTCTGATG		
-----+-----+-----+-----+-----+		
SerTrpLeuGlyAsnIleIleMetTyrAlaProThrLeuTrpAlaArgMetIleLeuMet		
1810 1820		

FIG. 5M

37/92

5470	5490	5510
ACTCACTTCTTCTCCATCCTCTAGCACAGGAGCAACTTGAAAAAGCCCTGGACTGCCAG		
-----+-----+-----+-----+-----+		
ThrHisPhePheSerIleLeuLeuAlaGlnGluGlnLeuGluLysAlaLeuAspCysGln		
1830		
1840		
5530	5550	5570
ATCTACGGGCCTGTTACTCCATTGAGCCACTGACCTACCTCAGATCATTGAACGACTC		
-----+-----+-----+-----+-----+		
IleTyrGlyAlaCysTyrSerIleGluProLeuAspLeuProGlnIleIleGluArgLeu		
1850		
1860		
5590	5610	5630
CATGGCCTTAGCGCATTTCACTCCATAGTTACTCTCAGGTGAGATCAATAGGGTGGCT		
-----+-----+-----+-----+-----+		
HisGlyLeuSerAlaPheSerLeuHisSerTyrSerProGlyGluIleAsnArgValAla		
1870		
1880		
5650	5670	5690
TCATGCCTCAGGAAACTGGGTACCACCCCTGCGAGTCTGGAGACATGGGCCAGGAGC		
-----+-----+-----+-----+-----+		
SerCysLeuArgLysLeuGlyValProProLeuArgValTrpArgHisArgAlaArgSer		
1890		
1900		
5710	5730	5750
GTCCCGCGTAGGCTACTGTCCCAGGGGGGGAGGGCCGCACCTGTGGCAAGTACCTCTC		
-----+-----+-----+-----+-----+		
ValArgAlaArgLeuLeuSerGlnGlyGlyArgAlaAlaThrCysGlyLysTyrLeuPhe		
1910		
1920		
5770	5790	5810
AACTGGGCAGTGAAGACCAAACCTAAACACTCCAATCCGGCTGCGTCCCAGCTGGAC		
-----+-----+-----+-----+-----+		
AsnTrpAlaValLysThrLysLeuLysLeuThrProIleProAlaAlaSerGlnLeuAsp		
1930		
1940		
5830	5850	5870
TTGTCCGGCTGGTTCGTTGCTGGTACAGCGGGGAGACATATATCACAGCCTGTCTCGT		
-----+-----+-----+-----+-----+		
LeuSerGlyTrpPheValAlaGlyTyrSerGlyGlyAspIleTyrHisSerLeuSerArg		
1950		
1960		

FIG. 5N

38/92

5890 5910 5930
GCCCGACCCCGCTGGTTCATGCTGTGCCTACTCCTACTTCTGTAGGGTAGGCATCTAC
-----+-----+-----+-----+-----+
AlaArgProArgTrpPheMetLeuCysLeuLeuLeuSerValGlyValGlyIleTyr
 1970 1980

5950 5955
CTGCTCCCCAACCGA (SEQ. ID. NO. 5)
-----+-----
LeuLeuProAsnArg (SEQ. ID. NO. 6)
 1985

FIG. 5O

1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
51 GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCCG
101 TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCAGGGCTGG CTTAACTATG
151 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
201 CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA
251 TTGCATACGT TGATCCATA TCATAATATG TACATTTATA TTGGCTCATG
301 TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT
351 AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT
401 ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG
451 CCCATTGACG TCAATAATGA CGTATGTTCC CATACTAACG CCAATAGGGA
501 CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG
551 GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
601 TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG
651 ACTTTCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG
701 GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTGACTC
751 ACGGGGATTG CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTGTTT
801 GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCCA
851 TTGACGCAGA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
901 AGCTCGTTA GTGAACCGTC AGATGCCCTG GAGACGCCAT CCACGCTGTT
951 TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA
1001 CGGTGCATTG GAACCGGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC
1051 CTATAGACTC TATAGGCACA CCCCTTGGC TCTTATGCAT GCTATACTGT
1101 TTTTGGCTTG GGGCCTATAC ACCCCCGCTT CCTTATGCTA TAGGTGATGG
1151 TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
1201 TATTGGTGAC GATACTTCC ATTACTAATC CATAACATGG CTCTTGCCA
1251 CAACTATCTC TATTGGCTAT ATGCCAATAC TCTGCTCTTC AGAGACTGAC
1301 ACGGACTCTG TATTTTACA GGATGGGTC CCATTATTAA TTTACAAATT
1351 CACATATACA ACAACGCCGT CCCCCGTGCC CGCAGTTTT ATTAAACATA
1401 GCGTGGGATC TCCACCGCAA TCTCGGGTAC GTGTTCCGGA CATGGGCTCT
1451 TCTCCGGTAG CGCGGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCCTC
1501 AGCGGCTCAT GGTGCGCTCGG CAGCTCCTTG CTCTAACAG TGGAGGCCAG
1551 ACTTAGGCAC AGCACAATGC CCACCACAC CAGTGTGCCG CACAAGGCCG
1601 TGGCGTAGG GTATGTGTCT GAAAATGAGC GTGGAGATTG GGCTCGCACG
1651 GCTGACGCAG ATGGAAGACT TAAGGCAGCG GCAGAAGAAG ATGCAGGCAG
1701 CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACCTCC GTTGCAGGTGC
1751 TGTTAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGCG
1801 CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTCCAT
1851 GGGTCTTTTC TGCAAGTCACC GTCCTTAGAT CTAGGTACCA GATATCAGAA
1901 TTCAGTCGAC AGCGGCCGCG ATCTGCTGTG CCTCTAGTT GCCAGCCATC
1951 TGTTGTTGC CCCTCCCCCG TGCTCCCTT GACCCCTGGAA GGTGCCACTC
2001 CCACTGTCCT TTCCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT
2051 AGGTGTCATT CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGGA

40/92

2101 GGATTGGGAA GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG
2151 CCGCTGCGGC CAGGTGCTGA AGAATTGACC CGGTTCCCTCC TGGGCCAGAA
2201 AGAACGAGC ACATCCCCTT CTCTGTGACA CACCCGTCC ACGCCCCCTGG
2251 TTCTTAGTTC CAGCCCCACT CATAGGACAC TCATAGCTCA GGAGGGCTCC
2301 GCCTTCAATC CCACCCGCTA AAGTACTTGG AGCGGTCTCT CCCTCCCTCA
2351 TCAGCCCACC AAACCAAACC TAGCCTCCAA GAGTGGGAAG AAATTAAAGC
2401 AAGATAGGCT ATTAAGTGCA GAGGGAGAGA AAATGCCTCC AACATGTGAG
2451 GAAGTAATGA GAGAAATCAT AGAATTCTT CCGCTTCCTC GCTCACTGAC
2501 TCGCTGCGCT CGGTCGTTCG GCTGCGGCGA GCGGTATCAG CTCACTCAA
2551 GGCAGTAATA CGGTTATCCA CAGAACAGG GGATAACGCA GGAAAGAAC
2601 TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA GGCCGCGTTG
2651 CTGGCGTTT TCCATAGGCT CCGCCCCCT GACGAGCATC ACAAAAATCG
2701 ACGCTCAAGT CAGAGGTGGC GAAACCCGAC AGGACTATAA AGATACCAGG
2751 CGTTTCCCCC TGGAAGCTCC CTCGTGCGCT CTCCTGTTCC GACCCTGCG
2801 CTTACCGGAT ACCTGTCCGC CTTTCTCCCT TCGGGAAAGCG TGGCGCTTTC
2851 TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC GTTCGCTCCA
2901 AGCTGGGCTG TGTGCACGAA CCCCCCGTTC AGCCCGACCG CTGCGCCTTA
2951 TCCGGTAACT ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC
3001 ACTGGCAGCA GCCACTGGTA ACAGGATTAG CAGAGCGAGG TATGTAGGCG
3051 GTGCTACAGA GTTCTTGAAG TGGTGGCCTA ACTACGGCTA CACTAGAAGA
3101 ACAGTATTG GTATCTGCGC TCTGCTGAAG CCAGTTACCT TCGGAAAAAG
3151 AGTTGGTAGC TCTTGATCCG GCAAACAAAC CACCGCTGGT AGCGGTGGTT
3201 TTTTGTTTG CAAGCAGCAG ATTACCGCAG GAAAAAAAGG ATCTCAAGAA
3251 GATCCTTGTGA TCTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAATC
3301 ACGTTAAGGG ATTTGGTCA TGAGATTATC AAAAAGGATC TTCACCTAGA
3351 TCCTTTAAA TAAAAAAATGA AGTTTTAAAT CAATCTAAAG TATATATGAG
3401 TAAACTTGGT CTGACAGTTA CCAATGCTTA ATCAGTGAGG CACCTATCTC
3451 AGCGATCTGT CTATTCGTT CATCCATAGT TGCGTACTC GGGGGGGGG
3501 GGCCTGAGG TCTGCCTCGT GAAGAAGGTG TTGCTGACTC ATACCAGGCC
3551 TGAATCGCCC CATCATCCAG CCAGAAAGTG AGGGAGCCAC GGTTGATGAG
3601 AGCTTGTG TAGGTGGACC AGTTGGTGTAT TTTGAACCTT TGCTTGCCTA
3651 CGGAACGGTC TCGTTGTCG GGAAGATGCG TGATCTGATC CTTCAACTCA
3701 GCAAAAGTTC GATTTATTCA ACAAAAGCCGC CGTCCCGTCA AGTCAGCGTA
3751 ATGCTCTGCC AGTGTACAA CCAATTAACC AATTCTGATT AGAAAAACTC
3801 ATCGAGCATE AAATGAAAAT GCAATTATTAT CATATCAGGA TTATCAATAC
3851 CATATTTG AAAAAGCCGT TTCTGTAATG AAGGAGAAAA CTCACCGAGG
3901 CAGTTCCATA GGATGGCAAG ATCCTGGTAT CGGTCTGCGA TTCCGACTCG
3951 TCCAACATCA ATACAACCTA TTAATTCCC CTCGTAAAA ATAAGGTTAT
4001 CAAGTGGAGAA ATCACCATGA GTGACGACTG AATCCGGTGA GAATGGCAAA
4051 AGCTTATGCA TTTCTTCCA GACTTGTCA ACAGGCCAGC CATTACGCTC
4101 GTCATCAAAA TCACTCGCAT CAACCAAACC GTTATTTCATT CGTGATTGCG
4151 CCTGAGCGAG ACGAAATACCG CGATCGCTGT TAAAAGGACA ATTACAAACA

FIG. 6B

41/92

4201 GGAATCGAAT GCAACCGGCG CAGGAACACT GCCAGCGCAT CAACAATATT
4251 TTCACCTGAA TCAGGATATT CTTCTAACAC CTGGAATGCT GTTTTCCCGG
4301 GGATCGCAGT GGTGAGTAAC CATGCATCAT CAGGAGTACG GATAAAATGC
4351 TTGATGGTCG GAAGAGGCAT AAATTCCGTC AGCCAGTTA GTCTGACCAT
4401 CTCATCTGTA ACATCATTGG CAACGCTACC TTTGCCATGT TTCAGAAACA
4451 ACTCTGGCGC ATCGGGCTTC CCATACAATC GATAGATTGT CGCACCTGAT
4501 TGCCCCACAT TATCGCGAGC CCATTTATAC CCATATAAAT CAGCATCCAT
4551 GTTGGAAATT AATCGCGGCC TCGAGCAAGA CGTTTCCCGT TGAATATGGC
4601 TCATAAACACC CCTTGATTTA CTGTTTATGT AAGCAGACAG TTTTATTGTT
4651 CATGATGATA TATTTTTATC TTGTGCAATG TAACATCAGA GATTTTGAGA
4701 CACAACGTGG CTTTCCCCCCC CCCCCCATTA TTGAAGCATT TATCAGGGTT
4751 ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA
4801 ATAGGGTTTC CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTCTAAGA
4851 AACCAATTATT ATCATGACAT TAACCTATAA AAATAGGCGT ATCACGAGGC
4901 CCTTTCGTC

FIG. 6C

42/92

1 CATCATCAAT AATATAACCTT ATTTTGGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT
 61 TTGTGACGTG CCCCGGGGCG TGGGAACGGG CGGGGTGACG TAGTAGTGTG GCAGAAAGTGT
 121 GATGTTGTA GTGTGGCGGA ACACATGTAA GCGCCGGATG TGGTAAAAGT GACGTTTTG
 181 GTGTGCAGCG GTGTACACGG GAAGTGACAA TTTTCGCGCG GTTTTAGGCG GATGTTGTAG
 241 TAAATTTGGG CGTAACCAAG TAATATTGG CCATTTCGC GGGAAAAGT AATAAGAGGA
 301 AGTGAATCT GAATAATTCT GTGTTACTCA TAGCGCGTAA TATTTGTCTA GGGCCGCGGG
 361 GACTTTGACC GTTTACGTGG AGACTCGCCC AGGTGTTTT CTCAGGTGTT TTCCGCGTTC
 421 CGGGTCAAAG TTGGCGTTT ATTATTATAG TCAGCTGACG CGCAGTGTAT TTATACCCGG
 481 TGAGTTCTC AAGAGGCCAC TCTTGAGTGC CAGCGAGTAG AGTTTCTCC TCCGAGCCGC
 541 TCCGACACCG GGACTGAAA TGAGACATAT TATCTGCCAC GGAGGTGTTA TTACCGAAGA
 601 AATGGCCGCC AGTCTTTGG ACCAGCTGAT CGAAGAGGTA CTGGCTGATA ATCTCCACC
 661 TCCTAGCCAT TTTGAACCAC CTACCCCTCA CGAACGTGTAT GATTTAGACG TGACGGCCCC
 721 CGAAGATCCC AACGAGGAGG CGGTTTCGCA GATTTTCCC GAGTCTGTAA TGTTGGCGGT
 781 GCAGGAAGGG ATTGACTTAT TCACTTTCC GCCGGCGCCC GGTTCTCCGG AGCCGCCCTCA
 841 CCTTTCCCG CAGCCCGAGC AGCCGGAGCA GAGAGCCTTG GGTCCGGTTT CTATGCCAAA
 901 CCTTGTGCCG GAGGTGATCG ATCTTACCTG CCACCGAGGCT GGCTTCCAC CCAGTGACGA
 961 CGAGGATGAA GAGGGTGAGG AGTTTGTGTT AGATTATGTG GAGCACCCCG GGCACGGTTG
 1021 CAGGTCCTGT CATTATCACC GGAGGAATAC GGGGGACCCA GATATTATGT GTTCGCTTTG
 1081 CTATATGAGG ACCTGTGGCA TGTTTGTCTA CAGTAAGTGA AAAATTATGG GCAGTGGGTG
 1141 ATAGAGTGGT GGGTTTGGTG TGGTAATT TTGTTTAATT TTTACAGTT TGTGGTTAA
 1201 AGAATTGGT ATTGTGATTT TTGAAAGGT CCTGTGTCTG AACCTGAGCC TGAGCCGAG
 1261 CCAGAACCGG AGCCTGCAAG ACCTACCCGG CGTCCTAAAT TGGTGCCTGC TATCCTGAGA
 1321 CGCCCGACAT CACCTGTGTC TAGAGAATGC AATAGTAGTA CGGATAGCTG TGACTCCGGT
 1381 CCTTCTAACCA CACCTCCTGA GATACACCCG GTGGTCCCGC TGTGCCCAT TAAACCAGTT
 1441 GCGTGTGAGAG TTGGTGGCGC TCGCCAGGCT GTGGAAATGTA TCGAGGACTT GCTTAACGAG
 1501 TCTGGCAAC CTTGGACTT GAGCTGTAAA CGCCCCAGGC CATAAGGTGT AAACCTGTGA
 1561 TTGCGTGTGT GGTAAACGCC TTGTTTGCT GAATGAGTTG ATGTAAGTTT AATAAAGGGT
 1621 GAGATAATGT TAACTTGCA TGGCGTGTAA ATGGGGCGG GGCTTAAAGG GTATATAATG
 1681 CGCCGTGGGC TAATCTGGT TACATCTGAC CTCATGGAGG CTTGGGAGTG TTTGGAAGAT
 1741 TTTCTGCTG TGCCTAACTT GCTGGAACAG AGCTCTAACCA GTACCTCTG GTTTGGAGG
 1801 TTTCTGTGGG GCTCCTCCCA GGCAAAGTTA GTCTGCAGAA TTAAGGAGGA TTACAAGTGG
 1861 GAATTGAAAG AGCTTTGAA ATCCTGTGGT GAGCTGTGTTG ATTCTTGAA TCTGGGTAC
 1921 CAGGGCCTTT TCCAAGAGAA GGTCTCAAG ACTTTGGATT TTTCCACACC GGGGCGCGCT
 1981 GCGGCTGCTG TTGCTTTTTT GAGTTTATA AAGGATAAAAT GGAGCGAAGA AACCCATCTG
 2041 AGCGGGGGGT ACCTGTGGAA TTTCTGGCC ATGCATCTGT GGAGAGCGGT GGTGAGACAC
 2101 AAGAATCGCC TGCTACTGTT GTCTTCCGTC CGCCCGGCAA TAATACCGAC GGAGGACCAA
 2161 CAGCAGGAGG AAGCCAGGCG CGGGCGCCGG CAGGAGCAGA GCCCATGGAA CCCGAGAGCC
 2221 GGCCTGGACC CTCGGGAATG AATGTTGTAC AGGTGGCTGA ACTGTTCCA GAACTGAGAC
 2281 GCATTTAAC CATTAAACGAG GATGGGGCAGG GGCTAAAGGG GGTAAAGAAG GAGGGGGGG
 2341 CTTCTGAGGC TACAGAGGAG GCTAGGAATC TAACTTTAG CTTAATGACC AGACACCGTC
 2401 CTGAGTGTGT TACTTTTCAG CAGATTAAGG ATAATTGCGC TAATGAGCTT GATCTGCTGG
 2461 CGCAGAAGTA TTCCATAGAG CAGCTGACCA CTTACTGGCT GCAGCCAGGG GATGATTTG

FIG. 7A

43/92

2521 AGGAGGCTAT TAGGGTATAT GCAAAGGTGG CACTTAGGCC AGATTGCAAG TACAAGATTA
 2581 GCAAACCTGT AAATATCAGG AATTGTTGCT ACATTTCTGG GAACGGGGCC GAGGTGGAGA
 2641 TAGATAACGGA GGATAGGGTG GCCTTAGAT GTAGCATGAT AAATATGTGG CGGGGGGTGC
 2701 TTGGCATGGA CGGGTGGTT ATTATGAATG TGAGGTTAC TGGTCCAAT TTAGCGGTA
 2761 CGGTTTCCT GGCCAATACC AATCTTATCC TACACGGTGT AAGCTTCTAT GGGTTAACAA
 2821 ATACCTGTGT GGAAGCCTGG ACCGATGTA GGTTCGGGG CTGTGCCTT TACTGCTGCT
 2881 GGAAGGGGT GGTGTGTCGC CCCAAAAGCA GGGCTCAAT TAAGAAATGC CTGTTGAAA
 2941 GGTGTACCTT GGGTATCCTG TCTGAGGGTA ACTCCAGGGT GCGCCACAAT GTGGCCTCCG
 3001 ACTGTGGTTG CTTTATGCTA GTGAAAAGCG TGGCTGTGAT TAAGCATAAC ATGGTGTGTC
 3061 GCAACTGCGA GGACAGGGCC TCTCAGATGC TGACCTGCTC GGACGGCAAC TGTCACTTGC
 3121 TGAAGACCAT TCACGTAGCC AGCCACTCTC GCAAGGCCTG GCCAGTGTGTT GAGCACAA
 3181 TACTGACCCG CTGTTCCCTG CATTGGGTA ACAGGGAGGGG GGTGTTCCCTA CCTTACCAAT
 3241 GCAATTGAG TCACACTAAG ATATTGCTTG AGCCCAGAG CATGTCCAAG GTGAACCTGA
 3301 ACGGGGTGTG TGACATGACC ATGAAGATCT GGAAGGTGCT GAGGTACGAT GAGACCCGCA
 3361 CCAGGTGCG ACCCTGCGAG TGTGGGGTA AACATATTAG GAACCAGCCT GTGATGCTGG
 3421 ATGTGACCGA GGAGCTGAGG CCCGATCACT TGGTGCTGGC CTGCACCCGC GCTGAGTTG
 3481 GCTCTAGCGA TGAAGATACA GATTGAGGTG CTGAAATGTG TGGGCGTGGC TTAAGGGTGG
 3541 GAAAGAATAT ATAAGGTGGG GGTCTCATGT AGTTTGTAT CTGTTTGCA GCAGCCGCCG
 3601 CCATGAGCGC CAACTCGTT GATGGAAGCA TTGTGAGCTC ATATTGACA ACGCGCATGC
 3661 CCCCCATGGGC CGGGGTGCGT CAGAATGTGA TGGGCTCCAG CATTGATGGT CGCCCCGTCC
 3721 TGCCCGAAA CTCTACTACC TTGACCTACG AGACCGTGTG TGGAACGCCG TTGGAGACTG
 3781 CAGCCTCCGC CGCCGCTTCA GCGCTGCAG CCACCGCCCG CGGGATTGTG ACTGACTTTG
 3841 CTTTCCTGAG CCCGCTTGCA AGCAGTCAG CTTCCCGTTC ATCCGCCGC GATGACAAGT
 3901 TGACGGCTCT TTTGGCACAA TTGGATTCTT TGACCCGGGA ACTTAATGTC GTTCTCAGC
 3961 AGCTGTTGGA TCTGCGCCAG CAGGTTCTG CCCTGAAGGC TTCCCTCCCT CCCAATGCGG
 4021 TTTAAAACAT AAATAAAAAC CAGACTCTGT TTGGATTGGG ATCAAGCAAG TGTCTTGCTG
 4081 TCTTTATTAA GGGGTTTGC GCGCGGGTA GGCCCGGGAC CAGCGGTCTC GGTGTTGAG
 4141 GGTCTGTGT ATTTTTCCA GGACGTGGTA AAGGTGACTC TGGATGTTCA GATACATGGG
 4201 CATAAGCCCG TCTCTGGGGT GGAGGTAGCA CCACTGCAGA GCTTCATGCT CGGGGGTGGT
 4261 GTTGTAGATG ATCCAGTCGT AGCAGGAGCG CTGGCGTGG TGCTAAAAA TGTCTTCAG
 4321 TAGCAAGCTG ATTGCCAGGG GCAGGCCCTT GGTGTAAGTG TTTACAAAGC GGTAAAGCTG
 4381 GGATGGGTGC ATACGTGGGG ATATGAGATG CATCTTGGAC TGTATTTTA GGTTGGCTAT
 4441 GTTCCCAGCC ATATCCCTCC GGGGATTCA GTTGTGCAGA ACCACCAGCA CAGTGTATCC
 4501 GGTGCACTTG GGAAATTGT CATGTAGCTT AGAAGGAAAT GCGTGGAAAGA ACTTGGAGAC
 4561 GCCCTTGTGA CCTCCAAGAT TTTCCATGCA TTCGTCCATA ATGATGGCAA TGGGCCACG
 4621 GGCAGCGGCC TGGGCGAAGA TATTCTGGG ATCACTAACG TCATAGTTGT GTTCCAGGAT
 4681 GAGATCGTCA TAGGCCATT TTACAAAGCG CGGGCGGAGG GTGCCAGACT GCGGTATAAT
 4741 GGTTCCATCC GGCCCAAGGG CGTAGTTACC CTCACAGATT TGCATTTCCC ACGCTTGAG
 4801 TTCAGATGGG GGGATCATGT CTACCTGCGG GGCGATGAAG AAAACCGTTT CGGGGGTAGG
 4861 GGAGATCAGC TGGGAAGAAA GCAGGTTCCCT AAGCAGCTGC GACTTACCGC AGCCGGTGGG
 4921 CCCGTAAATC ACACCTATTA CCGGCTGCAA CTGGTAGTTA AGAGAGCTGC AGCTGCCGTC
 4981 ATCCCTGAGC AGGGGGGCCA CTTCGTTAAG CATGTCCCTG ACTTGCATGT TTTCCCTGAC

FIG. 7B

44/92

5041 CAAATCCGCC AGAAGGCAGT CGCCGCCAG CGATAGCAGT TCTTGCAAGG AAGCAAAGTT
 5101 TTTCAACGGT TTGAGGCCGT CCGCCGTAGG CATGCTTTG AGCGTTTGAC CAAGCAGTTC
 5161 CAGGCGGTCC CACAGCTCGG TCACGTGCTC TACGGCATCT CGATCCAGCA TATCTCCTCG
 5221 TTTCGCGGGT TGGGGCGGCT TTCGCTGTAC GGCAGTAGTC GGTGCTCGTC CAGACGGGCC
 5281 AGGGTCATGT CTTTCCACGG GCGCAGGGTC CTCGTCAGCG TAGTCTGGGT CACGGTGAAG
 5341 GGGTGCCTC CGGGGTGCAG GCTGGCCAGG GTGCGCTTGA GGCTGGTCCT GCTGGTGCTG
 5401 AAGCGCTGCC GGTCTTCGCC CTGCGCTCG GCCAGGTAGC ATTTGACCAT GGTGTCATAG
 5461 TCCAGCCCCC CCGCGGCCGTG GCCCTTGGCG CGCAGCTTGC CCTTGGAGGA GGCGCCGCAC
 5521 GAGGGGCAGT GCAGACTTTT AAGGGCGTAG AGCTTGGCG CGAGAAATAC CGATTCCGGG
 5581 GAGTAGGCAT CCGCGCCGCA GGCCCCGCAG ACGGTCTCGC ATTCCACGAG CCAGGTGAGC
 5641 TCTGGCCGTT CGGGGTCAAA AACCAAGGTTT CCCCCATGCT TTTTGATGCG TTTCTTACCT
 5701 CTGGTTCCA TGAGCCGGTG TCCACGCTCG GTGACGAAA GGCTGTCCGT GTCCCCGTAT
 5761 ACAGACTTGA GAGGCCCTGTC CTCGAGCGGT GTTCCCGGGT CCTCCTCGTA TAGAAACTCG
 5821 GACCACCTCG AGACGAAGGC TCGCGTCCAG GCCAGCACGA AGGAGGCTAA GTGGGAGGGG
 5881 TAGCGGTCGT TGTCCACTAG GGGGTCCACT CGCTCCAGGG TGTGAAGACA CATGTCGCC
 5941 TCTTCGGCAT CAAGGAAGGT GATTGGTTTA TAGGTGTAGG CCACGTGACC GGGTGTTCCT
 6001 GAAGGGGGGC TATAAAAGGG GGTGGGGCG CGTTCGTCCT CACTCTCTTC CGCATCGCTG
 6061 TCTGCGAGGG CCAGCTGTTG GGGTGAGTAC TCCCTCTCAA AAGCGGGCAT GACTTCTGCG
 6121 CTAAGATTGT CAGTTTCAA AAACGAGGGAG GATTGATAT TCACCTGGCC CGCGGTGATG
 6181 CCTTTGAGGG TGAGCCGCGTC CATCTGGTCA GAAAAGACAA TCTTTTTGTT GTCAAGCTTG
 6241 GTGGCAAACG ACCCGTAGAG GGCGTTGGAC AGCAACTTGG CGATGGAGGG CAGGGTTGG
 6301 TTTTGTCGC GATCGGCGCG CTCCCTGGCC GCGATGTTA GCTGCACGTA TTGCGCGCGA
 6361 ACGCACCGCC ATTCTGGAAA GACGGTGGTG CGCTCGTCGG GCACTAGGTG CACGCGCCAA
 6421 CGCGGGTTGTG GCAGGGTGAC AAGGTCAACG CTGGTGGCTA CCTCTCCCGCG TAGGCGCTCG
 6481 TTGGTCCAGC AGAGGGGGCC GCCCTTGGCG GAGCAGAATG GCGGTAGTGG GTCTAGCTGC
 6541 GTCTCGTCGG GGGGGTCTGC GTCCACGGTA AAGACCCGG GCAGCAGGGCG CGCGTCGAAG
 6601 TAGTCTATCT TGCACTCCTTG CAAGTCTAGC GCCTGCTGCC ATGCGCGGGC GGCAAGCGCG
 6661 CGCTCGTATG GGGTGAATGG GGGACCCAT GGCATGGGGT GGGTGAGGGC GGAGGCGTAC
 6721 ATGCCGCAA TGTCGTAAC GTAGAGGGC TCTCTGAGTA TTCCAAGATA TGTAGGGTAG
 6781 CATCTTCCAC CGCGGATGCT GGCGCGCACG TAATCGTATA GTTCGTGCGA GGGAGCGAGG
 6841 AGGTCCGGAC CGAGGGTTGCT ACGGGGGGC TGCTCTGCTC GGAAGACTAT CTGCTGAAAG
 6901 ATGGCATGTG AGTTGGATGA TATGGTTGGA CGCTGGAAGA CGTTGAAGCT GGCGTCTGTG
 6961 AGACCTACCG CGTCACGCAC GAAGGAGGGC TAGGAGTCGC GCAGCTTGGT GACCAGCTCG
 7021 GCGGTGACCT GCACGTCTAG GGCGCAGTAG TCCAGGGTT CCTTGATGAT GTCATACTTA
 7081 TCCTGTCCT TTTTTTCCA CAGCTCGCG TTGAGGACAA ACTCTTCGCC GTCTTCCAG
 7141 TACTCTTGGT TCGGAAACCC GTCGGCCCTCC GAACGGTAAG AGCCTAGCAT GTAGAACTGG
 7201 TTGACGGCCT GGTAGGGCGCA GCATCCCTTT TCTACGGGTA GCGCGTATGC CTGCGCGGCC
 7261 TTCCGGAGCG AGGTGTGGGT GAGCGCAAAG GTGCCCCCAA CCATGACTTT GAGGTACTGG
 7321 TATTGAAAGT CAGTGTGCTC GCATCCGCC TGCTCCCAGA GCAAAAAGTC CGTGCCTTT
 7381 TTGGAACGCG GGTTTGGCAG GGCGAAGGTG ACATCGTTGA AGAGTATCTT TCCCGCGCGA
 7441 GGCATAAAAGT TGCCTGTGAT GCGGAAGGGT CCCGGCACCT CGGAACGGTT GTTAATTACC
 7501 TGGCGGGCGA GCACGATCTC GTCAAAGCCG TTGATGTTGT GGCCCCACAAT GTAAAGTTCC

FIG. 7C

45/92

7561 AAGAACGCGC GGATGCCCTT GATGGAAGGC AATTTCCTAA GTTCCTCGTA GGTGAGCTCT
 7621 TCAGGGGAGC TGAGCCCGTG CTCTGAAAGG GCCCAGTCTG CAAGATGAGG GTTGGAAAGCG
 7681 ACGAATGAGC TCCACAGGTC ACGGGCCATT AGCATTGCA GGTGGTCGCG AAAGGTCTA
 7741 AACTGGCGAC CTATGCCAT TTTTCTGGG GTGATGCAGT AGAAGGTAAG CGGGTCTTGT
 7801 TCCCAGCGGT CCCATCCAAG GTCCCGGGCT AGGTCTCGCG CGGGGTCAC TAGAGGCTCA
 7861 TCTCCGCCGA ACTTCATGAC CAGCATGAAG GGCACGAGCT GCTTCCCCAA GGCCCCCATC
 7921 CAAGTATAGG TCTCTACATC GTAGGTGACA AAGAGACGCT CGGTGGAGG ATGGGAGCCG
 7981 ATCGGGAAAGA ACTGGATCTC CCGCCACCAG TTGGAGGAGT GGCTGTTGAT GTGGTGAAAG
 8041 TAGAAGTCCC TGGCACGGGC CGAACACTCG TGCTGGCTTT TGTAACACG TGCGCAGTAC
 8101 TGGCAGCGGT GCACGGGCTG TACATCTGC ACGAGGTTGA CCTGACGACC GGCACAAAGG
 8161 AAGCAGAGTG GGAATTGAG CCCCTCGCCT GGCGGGTTTG GCTGGTGGTC TTCTACTTCG
 8221 GCTGCTGTC CTTGACCGTC TGGCTGCTG AGGGGAGTTA CGGTGGATCG GACCACACG
 8281 CGCGCGAGC CCAAAGTCCA GATGTCGCG CGCGCGGTC GGAGCTTGAT GACAACATCG
 8341 CGCAGATGGG AGCTGTCCAT GGTCTGGAGC TCCCGCGGCG TCAGGTCAGG CGGGAGCTCC
 8401 TGCAGGTTA CCTCGCATAG CCGGGTCAGG GCGCGGGCTA GGTCCAGGTG ATACCTGATT
 8461 TCCAGGGGCT GGTGGTGGC GGCGTCGATG CCTTGCAAGA GGCGCAGTC CGCGCGCGCG
 8521 ACTACGGTAC CGCGCGCGG GCGGTGGGCC GCGGGGGTGT CCTTGGATGA TGCATCTAAA
 8581 AGCGGTGACG CGGGCGGGCC CCCGGAGGTA GGGGGGGCTC GGGACCCGCC GGGAGAGGGG
 8641 GCAGGGGAC GTCGGCGCCG CGCGCGGGCA GGAGCTGGTG CTGCGCGCGG AGGTTGCTGG
 8701 CGAACGCGAC GACGCGCGG TTGATCTCCT GAATCTGGCG CCTCTGCGTG AAGACGACGG
 8761 GCCCAGGTGAG CTTGAACCTG AAAGAGAGTT CGACAGAAC AATTTCGGTG TCGTTGACGG
 8821 CGGCCTGGCG CAAAATCTCC TGCACGTCTC CTGAGTTGTC TTGATAGGCG ATCTCGGCCA
 8881 TGAACGTCTC GATCTCTTCC TCCTGGAGAT CTCCCGCTCC GGCTCGCTCC ACGGTGGCGG
 8941 CGAGGTCGTT GGAGATGCGG GCCATGAGCT GCGAGAAGGC GTTGAGGCCT CCCTCGTTCC
 9001 AGACGCGGCT GTAGACCACG CCCCCCTCGG CATCGCGGCC GCGCATGACC ACCTGCGCGA
 9061 GATTGAGCTC CACGTGCCGG GCGAAGACGG CGTAGTPTCG CAGGCCTGAAAGAGGTAGT
 9121 TGAGGGTGGT GGCGGTGTGT TCTGCCACGA AGAAAGTACAT AACCCAGCGC CGCAACGTGG
 9181 ATTGTTGAT ATCCCCCAAG GCCTCAAGGC GCTCCATGGC CTCGTAGAAG TCCACGGCGA
 9241 AGTTGAAAAA CTGGGAGTTG CGCGCCGACA CGGTTAACTC CTCCTCCAGA AGACGGATGA
 9301 GCTCGGGAC AGTGTGCGC ACCTCGCGCT CAAAGGCTAC AGGGGGCTCT TCTTCTTCTT
 9361 CAATCTCCTC TTCCATAAGG GCCTCCCCCTT CTTCTTCTTC TGGCGGGCGGT GGGGGAGGGG
 9421 GGACACGGCG GCGACGACGG CGCACCGGGA GGCGGTGAC AAAGCGCTCG ATCATCTCCC
 9481 CGCGGGACG GCGCATGGTC TCGGTGACGG CGCGGGCGTT CTCGCGGGGG CGCAGTTGGA
 9541 AGACGCCGCC CGTCATGTCC CGGTTATGGG TTGGCGGGGG GCTGCCGTGC GGCAGGGATA
 9601 CGGCCTAAC GATGCATCTC AACAAATTGTT GTGTAGGTAC TCCGCCACCG AGGGACCTGA
 9661 GCGAGTCCGC ATCGACCGGA TCGGAAAACC TCTCGAGAAA GGCGTCTAAC CAGTCACAGT
 9721 CGCAAGGTAG GCTGAGCACC GTGGCGGGCG GCAGCGGGCG GCGGTGGGG TTGTTCTGG
 9781 CGGAGGTGCT GCTGATGATG TAATTAAGT AGGCGGTCTT GAGACGGGG ATGGTCGACA
 9841 GAAGCACCAC TGCCTGGGT CGGGCTGCT GAATGCGCAG GCGGTGGCC ATGCCCCAGG
 9901 CTTCGTTTG ACATCGCGC AGGTCTTGT AGTAGTCTTG CATGAGCCTT TCTACCGCA
 9961 CTTCTCTTC TCCCTCTCT TGTCTGCAT CTCTGACAT TATCGCTGCG GCGGGCGCGG
 10021 AGTTGGCCG TAGGTGGCGC CCTCTTCCTC CCATGCGTGT GACCCGAAG CCCCTCATCG

FIG. 7D

10081 GCTGAAGCAG GGCCAGGTG GCGACAACGC GCTCGGCTAA TATGGCCTGC TGCACCTGCG
 10141 TGAGGGTAGA CTCGAAGTCG TCCATGTCCA CAAAGCGGTG GTATGCGCCC GTGTTGATGG
 10201 TGTAAGTGCA GTTGGCCATA ACGGACCAGT TAACGGTCTG GTGACCCGGC TCGGAGAGCT
 10261 CGGTGTACCT GAGACGCGAG TAAGCCCTG AGTCAAAGAC GTAGTCGTTG CAAGTCGCA
 10321 CCAGGTACTG GTATCCCACC AAAAAGTGC GCGGCGGCTG GCGGTAGAGG GGCCAGCGTA
 10381 GGGTGGCCGG GGCTCCGGG GCGAGGTCTT CCAACATAAG GCGATGATAT CCGTAGATGT
 10441 ACCTGGACAT CCAGGTGATG CCGGCGGCG TGTTGGAGGC GCGCGGAAAG TCACGGACGC
 10501 GGTTCCAGAT GTTGCAGC GGCAAAAAGT GCTCCATGGT CGGGACGCTC TGGCCGGTCA
 10561 GCGCGCGCA GTCGTTGACG CTCTAGACCG TGCAAAAGGA GAGCCTGTA GCGGGCACTC
 10621 TTCCGTGGTC TGTTGGATAA ATTGCGAAGG GTATCATGGC GGACGACCGG GGTTCGAAC
 10681 CGGATCCGG CCGTCCGCCG TGATCCATGC GGTTACCGCC CGCGTGTGCA ACCCAGGTGT
 10741 GCGACGTCAG ACAACGGGG AGCGCTCCTT TTGGCTCCT TCCAGGCGCG GCGGATGCTG
 10801 CGCTAGCTT TTTGGCCACT GGCCGCGCGC GGCCTAAGCG GTTACGGCTGG AAAGCGAAAG
 10861 CATTAAGTGG CTCGCTCCCT GTAGCCGGAG GGTTATTTTC CAAGGGTTGA GTCGCGGGAC
 10921 CCCCGGTTCG AGTCTCGGGC CGGCGGGACT GCGGCGAACG GGGGTTGCC TCCCCGTCA
 10981 GCAAGACCCC GCTTGCAAAT TCCTCCGAA ACAGGGACGA GCCCCTTTTTG TGCTTTCCC
 11041 AGATGCATCC GGTGCTGCGG CAGATGCGCC CCCCTCCTCA GCAGCGGCAA GAGCAAGAGC
 11101 AGCGGCAGAC ATGCAGGGCA CCCTCCCCCT CTCCCTACCGC GTCAGGAGGG GCAACATCCG
 11161 CGGCTGACGC GGGCGCAGAT GGTGATTACG AACCCCCCGCG GCGCCGGACCG CGGCACACT
 11221 TGGACTTGGA GGAGGGCGAG GGCTGGCGC GGCTAGGAGC GCCCTCTCCT GAGCGACACC
 11281 CAAGGGTGC GCTGAAGCGT GACACGGCG AGGCGTACGT GCCGCGGCAG AACCTGTTTC
 11341 GCGACCGCGA GGGAGAGGAG CCCGAGGAGA TGCGGGATCG AAAGTCCAT GCAGGGCGCG
 11401 AGTTGCGGCA TGGCCTGAAC CGCGAGCGGT TGCTGCGCA GGAGGACTTT GAGCCGACG
 11461 CGCGGACCGG GATTAGTCCC GCGCGCCAC ACGTGGCGC CGCCGACCTG GTAACCGCGT
 11521 ACGAGCAGAC GGTGAACCAAG GAGATTAAC TTCAAAAAAG CTTAACAAAC CACGTGCGCA
 11581 CGCTTGTGGC GCGCGAGGAG GTGGCTATAG GACTGATGCA TCTGTGGAC TTTGTAAGCG
 11641 CGCTGGAGCA AAACCCAAAT AGCAAGCCGC TCATGGCGCA GCTGTTCCCT ATAGTGCAGC
 11701 ACAGCAGGGA CAACGAGGCA TTCAGGGATG CGCTGCTAA CATAGTAGAG CCCGAGGGCC
 11761 GCTGGCTGCT CGATTTGATA AACATTCTGC AGAGCATAGT GGTGCAGGAG CGCAGCTTG
 11821 GCCTGGCTGA CAAGGTGGCC GCCATTAAC ATTCCATGCT CAGTCTGGC AAGTTTTACG
 11881 CCCGCAAGAT ATACCATACC CCTTACGTT CCATAGACAA GGAGGTAAAG ATCGAGGGGT
 11941 TCTACATGCG CATGGCGCTG AAGGTGCTTA CCTTGAGCGA CGACCTGGC GTTTATCGCA
 12001 ACGAGCGCAT CCACAAGGCC GTGAGCGTGA GCGGGCGCG CGAGCTCAGC GACCGCGAGC
 12061 TGATGCACAG CCTGCAAAGG GCCCTGGCTG GCACGGGCAG CGCGATAGA GAGGCCGAGT
 12121 CCTACTTGA CGCGGGCGCT GACCTGCGCT GGGCCCAAG CGACGCGCC CTGGAGGCAG
 12181 CTGGGCCGG ACCTGGGCTG GCGGTGGCAC CGCGCGCGC TGGCAACGTC GCGGGCGTGG
 12241 AGGAATATGA CGAGGACGAT GAGTACGAGC CAGAGGACGG CGAGTACTAA GCGGTGATGT
 12301 TTCTGATCAG ATGATGCAAG ACGCAACGGA CCCGGCGGTG CGGGCGGCC TGCAGAGCCA
 12361 GCGGTCCGGC CTTAACTCCA CGGACGACTG GCGCCAGGTC ATGGACCGCA TCATGTCGCT
 12421 GACTGCGCGC AACCTGACG CGTTCCGGCA GCAGCCGCAG GCCAACCGGC TCTCCGCAAT
 12481 TCTGGAAGCG GTGGTCCCGG CGCGCGAAA CCCCCACGCAC GAGAAGGTGC TGGCGATCGT
 12541 AAACGCGCTG GCCGAAAACA GGGCCATCCG GCCCGATGAG GCCGGCCTGG TCTACGACGC

47/92

12601 GCTGCTTCAG CGCGTGGCTC GTTACAACAG CAGCAACGTG CAGACCAACC TGGACCGGCT
12661 GGTGGGGAT GTGCGCGAGG CGGTGGCGA GCGTGAGCGC GCGCAGCAGC AGGGCAACCT
12721 GGGCTCCATG GTTGCACCAA ACGCCTCCCT GAGTACACAG CCCGCCAACG TGCCGCGGGG
12781 ACAGGAGGAC TACACCAACT TTGTGAGCGC ACTGCGGCTA ATGGTACTG AGACACCGCA
12841 AAGTGAGGTG TATCAGTCCG GGCCAGACTA TTTTTCCAG ACCAGTAGAC AAGGCCTGCA
12901 GACCGTAAAC CTGAGCCAGG CTTTCAAGAA CTTGCAGGGG CTGTGGGGGG TGCGGGCTCC
12961 CACAGGCGAC CGCGCGACCG TGTCTAGCTT GCTGACGCC AACTCGGCC TGTTGCTGCT
13021 GCTAAATAGCG CCCTTCACGG ACAGTGGCAG CGTGTCCCAG GACACATACC TAGGTCACTT
13081 GCTGACACTG TACCGCGAGG CCATAGGTCA GGCGCATGTG GACGAGCATA CTTTCCAGGA
13141 GATTACAAGT GTTAGCCCG CGCTGGGCA GGAGGACACG GGCAGCCTGG AGGCAACCC
13201 GAACTACCTG CTGACCAACC GGCGGCAAAA AATCCCCTCG TTGCACAGTT TAAACAGCGA
13261 GGAGGAGCGC ATTTCGCGCT ATGTGCAGCA GAGCGTGAGC CTTAACCTGA TGCGCGACGG
13321 GGTAAACGCC AGCGTGGCGC TGGACATGAC CGCGCGCAAC ATGGAACCGG GCATGTATGC
13381 CTCAAACCGG CCGTTTATCA ATCGCCTAAT GGACTACTTG CATCGCGCG CCGCCGTGAA
13441 CCCCGAGTAT TTCACCAATG CCATCTTGAA CCCGCACTGG CTACCGCCCC CTGGTTTCTA
13501 CACCGGGGAA TTGAGGTGC CCGAGGGTAA CGATGGATTG CTCTGGGACG ACATAGACGA
13561 CAGCGTGTTC TCCCCGCAAC CGCAGACCC GCTAGAGTTG CAACAAACGG AGCAGGGCAGA
13621 GCGGGCGCTG CGAAAGGAAA GCTTCCCGAG GCCAAGCAGC TTGTCCGATC TAGGCCTGCG
13681 GGCCCCCGGG TCAGATGCTA GTAGCCCCATT TCCAAGCTTG ATAGGGTCTC TTACCAAGCAC
13741 TCGCACCACCC CGCCCGCGCC TGCTGGCGA GGAGGAGTAC CTAAACAACT CGCTGCTGCA
13801 GCCGCAGCGC GAAAAGAACCG TGCCCTCCGGC GTTCCCAC AACGGGATAG AGAGCCTAGT
13861 GGACAAGATG AGTAGATGGA AGACGTATGC GCAGGAGCAC AGGGATGTGC CCGGCCCCGCG
13921 CCCGCCACC CGTCGTCAAA GGCACGACCG TCAGCGGGGT CTGGTGTGGG AGGACGATGAA
13981 CTCGGCAGAC GACAGCAGCG TCTTGGATTT GGGAGGGAGT GGCAACCCGT TTGCACACCT
14041 TCGCCCCAGG CTGGGGAGAA TGTTTAAAA AAAGCATGAT GCAAAATAAA AAACTCACCA
14101 AGGCCATGGC ACCGAGCGTT GGTTTCTTG TATTCCCCTT AGTATGCGGC GCGCGCGAT
14161 GTATGAGGAA GGTCTCCCTC CCTCCTACGA GAGCGTGGTG AGCGCGGGCG CAGTGGCGGC
14221 GCGCCTGGGT TCACCCCTCG ATGCTCCCT GGACCCCGG TTCTGTGCCTC CGCGGTACCT
14281 GCGGCCTACC GGGGGGAGAA ACAGCATCCG TTACTCTGAG TTGGCACCCCC TATTGACAC
14341 CACCCGTGTG TACCTTGTGG ACAACAAGTC AACGGATGTG GCATCCCTGA ACTACCAGAA
14401 CGACCACAGC AACTTTCTAA CCACGGTCAT TCAAAACAAT GACTACAGCC CGGGGGAGGC
14461 AAGCACACAG ACCATCAATC TTGACGACCG GTCCCACTGG GGCAGCGACCC TGAAAACCAT
14521 CCTGCATACC AACATGCCAA ATGTGAACGA GTTCATGTTT ACCAATAAGT TTAAGGCGCG
14581 GGTGATGGTG TCGCGCTCGC TTACTAAGGA CAAACAGGTG GAGCTGAAAT ACGAGTGGGT
14641 GGAGTTCAAG CTGCCCGAGG GCAACTACTC CGAGACCATG ACCATAGACC TTATGAACAA
14701 CGCGATCGTG GAGCACTACT TGAAAGTGGG CAGGCAGAAC GGGGTTCTGG AAAGCGACAT
14761 CGGGGTAAAG TTTGACACCC GCAACTTCAG ACTGGGTTT GACCCAGTCA CTGGTCTTGT
14821 CATGCCTGGG GTATATACAA ACGAAGCCTT CCATCCAGAC ATCATTTCG TGCCAGGATG
14881 CGGGGTGGAC TTCAACCCACA GCGCGCTGAG CAAACTGTG GGCATCCGCA AGCGGCAACC
14941 CTTCCAGGAG GGCTTTAGGA TCACCTACGA TGACCTGGAG GGTGGTAACA TTCCCGCACT
15001 GTTGGATGTG GACGCCTACC AGGCAAGCTT GAAAGATGAC ACCGAACAGG GCGGGGGTGG
15061 CGCAGGCGGC GGCAACAAACA GTGGCAGCGG CGCGGAAGAG AACTCCAACG CGGCAGCTGC

FIG. 7F

48/92

15121 GGCAATGCAG CCGGTGGAGG ACATGAACGA TCATGCCATT CGCGGCGACA CCTTTGCCAC
 15181 ACGGGCGGAG GAGAAGCGCG CTGAGGCCGA GGCAGCGGCC GAAGCTGCCG CCCCCGCTGC
 15241 GGAGGCTGCA CAACCCGAGG TCGAGAACCC TCAGAAAGAAA CCGGTGATTA AACCCCTGAC
 15301 AGAGGACAGC AAGAAACGCA GTTACAACT AATAAGCAAT GACAGCACCT TCACCCAGTA
 15361 CCGCAGCTGG TACCTTGCAT ACAACTACGG CGACCCCTAG GCCGGGATCC GCTCATGGAC
 15421 CCTGCTTGC ACTCCTGACG TAACCTGCGG CTCGGAGCAG GTATACTGGT CGTTGCCGA
 15481 CATGATGCAA GACCCCGTGA CCTTCCGTC CACCGGCCAG ATCAGCAACT TTCCGGTGGT
 15541 GGGCGCCGAG CTGTTGCCCG TGCACTCAA GAGCTTCTAC AACGACCAGG CCGTCTACTC
 15601 CCAGCTCATC CGCCAGTTA CCTCTCTGAC CCACGTGTTA AATCGCTTTC CCGAGAACCA
 15661 GATTTGGCG CGCCCGCCAG CCCCCACCAT CACCACCGTC AGTAAAACG TTCTGCTCT
 15721 CACAGATCAC GGGACGCTAC CGCTGCGCAA CAGCATCGGA GGAGTCCAGC GAGTGACCAT
 15781 TACTGACGCC AGACGCCGCA CCTGCCCTA CGTTACAAG GCCCTGGGCA TAGTCTGCC
 15841 GCGCGTCTA TCGAGCCGCA CTTTTTGAGC AAGCATGTCC ATCCTTATAT CGCCCAGCAA
 15901 TAACACAGGC TGGGGCCTGC GCTTCCAAAG CAAGATTTT GGCGGGGCCA AGAAGCGTC
 15961 CGACCAACAC CCAGTGCAG TGCGCGGCCA CTACCGCGCG CCCTGGGGCG CGCACAAACG
 16021 CGGCCGCACT GGGCGCACCA CGTCGATGA CGCCATCGAC CGGGTGGTGG AGGAGGCGCG
 16081 CAACTACACG CCCACGCCGC CGCCAGTGTG CACCGTGGAC CGGGCCATTG AGACCGTGGT
 16141 GCGCGGAGCC CGGCGCTACG CTAAAATGAA GAGACGGCGG AGGCGCTAG CACGTCGCCA
 16201 CCGCCGCCGA CCCGGCACTG CCGCCCAACG CGCGCGGGCG GCCCTGCTTA ACCGCGCACG
 16261 TCGCACCGC CGACGGGCGG CCATGCGAGC CGCTCGAAGG CTGGCCGGCG GTATTGTCAC
 16321 TGTGCCCCCCC AGGTCCAGGC GACGAGCGGC CGCCGAGCA GCCGCGGCCA TTAGTGCTAT
 16381 GACTCAGGGT CGCAGGGGCA ACGTGTACTG GGTGCGCGAC TCGGTTAGCG GCCTGCGCGT
 16441 GCCC GTGCCG ACCCGCCCCC CGCGCAACTA GATTGCAATA AAAAATACT TAGACTCGTA
 16501 CTGTTGTATG TATCCAGCGG CGCGCGCGC CATCGAAGCT ATGTCCAAGC GCAAATCAA
 16561 AGAAGAGATG CTCCAGGTCA TCGCGCCCGA GATCTATGGC CCCCCGAAGA AGGAAGAGCA
 16621 GGATTACAAG CCCC GAAAGC TAAAGCGGGT CAAAAAGAAA AAGAAAGATG ATGATGATGA
 16681 TGAACCTGAC GACGAGGTGG AACTGTTGCA CGCGACCGCG CCCAGGCGAC GGGTACAGTG
 16741 GAAAGGTGCA CGCGTAAGAC GTGTTTGCG ACCCGGCACC ACCGTAGTCT TTACGCCCGG
 16801 TGAGCGCTCC ACCCGCACCT ACAAGCGCGT GTATGATGAG GTGTACGGCG ACGAGGACCT
 16861 GCTTGAGCAG GCCAACGAGC GCCTCGGGGA GTTGCCTAC GGAAAGCGGC ATAAGGACAT
 16921 GCTGGCGTTG CGCGTGGACG AGGGCAACCC AACACCTAGC CTAAAGCCCG TGACACTGCA
 16981 GCAGGTGCTG CCCCGCTTG CACCGTCCGA AGAAAAGCGC GGCTAAAGC GCGAGTCTGG
 17041 TGACTTGGCA CCCACCGTGC AGCTGATGGT ACCCAAGCGT CAGCGACTGG AAGATGTCTT
 17101 GGAAAAAATG ACCGTGGAGC CTGGGCTGGA GCCCCGAGGTC CGCGTGCAGC CAATCAAGCA
 17161 GGTGGCACCG GGACTGGCG TGCAAGACCGT GGACGTTAGC ATACCCACCA CCAGTAGCAC
 17221 TAGTATTGCC ACTGCCACAG AGGGCATGGA GACACAAACG TCCCCGGTTG CCTCGCGGGT
 17281 GGCAGATGCC CGGGTGCAGG CGGCCGCTGC GGCGCGTCC AAGACCTCTA CGGAGGTGCA
 17341 AACGGACCCG TGGATGTTTC GTGTTTACG CCCCCGGCGT CCGCGCCGTT CAAGGAAGTA
 17401 CGGCGCCGCC AGCGCGCTAC TGCCCCGATA TGCCCTACAT CCTTCCATCG CGCCTACCCC
 17461 CGGCTATCGT GGCTACACCT ACCGGCCAG AAGACGAGCA ACTACCCGAC GCCGAACCC
 17521 CACTGAAACC CGCCGCCGCC GTCGCCGTG CCAGCCCCGTG CTGGCCCCGA TTTCCGTGCG
 17581 CAGGGTGGCT CGCGAAGGAG GCAGGACCCCT GGTGCTGCCA ACAGCGCGCT ACCACCCAG

FIG. 7G

17641 CATCGTTAA AAGCCGGTCT TTGTGGTCT TGCAGATATG GCCCTCACCT GCCGCCCTCCG
 17701 TTTCCCGGTG CCGGGATTCC GAGGAAGAAT GCACCGTAGG AGGGGCATGG CGGGCACCGG
 17761 CCTGACGGGC GGCATCGTC GTGCGCACCA CCGGCGGC GGCGCGTCGC ACCGTCGCAT
 17821 GCGCGCGGT ATCCTGCCCT TCCTTATTCC ACTGATCGCC GCGCGATIG GCGCCGTGCC
 17881 CGGAATTGCA TCCGTGGCCT TGCAGGGCAGA GAGACACTGA TTAAAAACAA GTTACATGTG
 17941 GAAAATCAA AATAAAAGTC TGGACTCTCA CGCTCGCTTG GTCCTGTAAC TATTTGTAG
 18001 AATGGAAGAC ATCAACTTTC CGTCACTGGC CCCGCGACAC GGCTCGGCC CGTTCATGGG
 18061 AAACTGGCAA GATATCGGC CAAGCAATAT GAGCGGTGGC GCCTTCAGCT GGGGCTCGCT
 18121 GTGGAGCGGC ATTAAAAATT TCGGTTCCGC CGTTAAGAAC TATGGCAGCA AAGCCTGGAA
 18181 CAGCAGCACA GGCCAGATGC TGAGGGACAA GTTGAAGAG CAAAATTCC AACAAAAGGT
 18241 GGTAGATGGC CTGGCCTCTG GCATTAGCGG GGTGGTGGAC CTGGCCAACC AGGCAGTGCA
 18301 AAATAAGATT AACAGTAAGC TTGATCCCCG CCCTCCCGTA GAGGAGCCTC CACCGGCCGT
 18361 CGAGACAGTG TCTCCAGAGG GGCAGGTGCA AAAGCGTCCG CGACCCGACA GGGAGAAC
 18421 TCTGGTGACG CAAATAGACG AGCCTCCCTC GTACGAGGAG GCACTAAAGC AAGGCCTGCC
 18481 CACCACCCGT CCCATCGCGC CCATGGCTAC CGGAGTGCTG GGCCAGCACA CACCGTAAC
 18541 GCTGGACCTG CCTCCCCCCC CCGACACCCCA GCAGAAACCT GTGCTGCCAG GCCCGTCCGC
 18601 CGTTGTTGTA ACCCGTCCTA GCGCGCGTC CCTGCGCCGC GCCGCCAGCG GTCCCGGATC
 18661 GTTGCGGCCCT GTAGCCAGTG GCAACTGGCA AAGCACACTG AACAGCATCG TGGGTTTGGG
 18721 GGTGCAATCC CTGAAGCGCC GACGATGCTT CTGATAGCTA ACGTGTCGTA TGTGTCAT
 18781 GTATGCGTCC ATGTCGCCGC CAGAGGAGCT GCTGAGCCGC CGCGCAGCCG CTTTCCAAGA
 18841 TGGCTACCCC TTCGATGATG CCGCAGTGGT CTTACATGCA CATCTCGGGC CAGGACGCCT
 18901 CGGAGTACCT GAGCCCCGGG CTGGTGCAGT TCGCCCGCGC CACCGAGACG TACTTCAGCC
 18961 TGAATAACAA GTTTAGAAAC CCCACGGTGG CGCCTACGCA CGACGTGACC ACAGACCGGT
 19021 CTCAGCGTTT GACGCTGCGG TTCATCCCCG TGGACCGCGA GGATACTGCG TACTCGTACA
 19081 AGGCGCGGTT CACCTAGCT GTGGGTGATA ACCGTGTGCT AGACATGGCT TCCACGTACT
 19141 TTGACATCCG CGCGTGTGCT GACAGGGGCC CTACTTTAA GCCCTACTCT GGCACCGCCT
 19201 ACAACGCCTG GGGCCCCAAG GGTGGCCCCA ACTCGTGCAGA GTGGGAACAA AATGAAACTG
 19261 CACAAGTGGA TGCTCAAGAA CTTGACGAAG AGGAGAACAGA AGCCAATGAA GCTCAGGC
 19321 GAGAACAGGA ACAAGCTAAG AAAACCCATG TATATGCCCA GGCTCCACTG TCCGAAATAA
 19381 AAATAACTAA AGAAGGTCTA CAAATAGGAA CTGCGACGC CACAGTAGCA GGTGCCGGCA
 19441 AAGAAAATTT CGCAGACAAA ACTTTCAAC CTGAACCACA AGTAGGAGAA TCTCAATGGA
 19501 ACGAACCGGA TGCCACAGCA GCTGGTGGAA GGGTTCTTAA AAAGACAACT CCCATGAAAC
 19561 CCTGCTATGG CTCATACGCT AGACCCACCA ATTCCAACGG CGGACAGGGC GTTATGGTTG
 19621 AACAAAATGG TAAATTGGAA AGTCAAGTCG AAATGCAATT TTTTTCCACA TCCACAAATG
 19681 CCACAAATGA AGTTAACAT ATACAACCAA CAGTTGTATT GTACAGCGAA GATGAAAC
 19741 TGGAAACTCC AGATACTCAT CTTTCTTATA AACCTAAAAT GGGGGATAAA AATGCCAAAG
 19801 TCATGCTTGG ACAACAAGCA ATGCCAAACA GACCAATT CATTGCTTT AGAGACAATT
 19861 TTATTGGTCT CATGTATTAC AACAGCACAG GTAACATGGG TGTCTTGCT GGTCAAGGC
 19921 CGCAGTTGAA CGCTGTTGTA GATTGCAAG ACAGAAACAC AGAGCTGTCC TACCAAGCTT
 19981 TGCTTGATTG AATTGGCGAC AGAACAGAT ACTTTCAAT GTGGAATCAA GCTGTTGACA
 20041 GCTATGATCC AGATGTCAGA ATTATTGAGA ACCATGGAAC TGAGGATGAG TTGCCAAATT
 20101 ATTGCTTCC TCCTGGTGGAA ATTGGGATTA CTGACACTTT TCAAGCTGTT AAAACAACTG

50/92

20161 CTGCTAACGG GGACCAAGGC AATACTACCT GGCAAAAAGA TTCAACATTT GCAGAACGCA
 20221 ATGAAATAGG CGTGGGAAT AACTTGCCA TGGAAATTAA CCTGAATGCC AACCTATGGA
 20281 GAAATTTCCT TTACTCCAAT ATTGCGCTGT ACCTGCCAGA CAAGCTAAA TACAACCCCA
 20341 CCAATGTGGA AATATCTGAC AACCCCCAAC CCTACGACTA CATGAACAAG CGAGTGGTGG
 20401 CTCCTGGCCT TGTAGACTGC TACATTAACC TTGGGGCGCG CTGGTCTCTG GACTACATGG
 20461 ACAACGTTAA TCCCTTAAC CACCACCGCA ATGCGGGCCT GCGTTACCGC TCCATGTTGT
 20521 TGGGAAACGG CCGCTACGTG CCCTTCACA TTCAGGTGCC CCAAAAGTTT TTTGCCATTA
 20581 AAAACCTCCT CCTCCTGCCA GGCTCATACA CATATGAATG GAACTTCAGG AAGGATGTTA
 20641 ACATGGTTCT GCAGAGCTCT CTGGGAAACG ACCTTAGAGT TGACGGGGCT AGCATTAAAGT
 20701 TTGACAGCAT TTGTCTTAC GCCACCTCT TCCCCATGGC CCACAAACACG GCCTCCACGC
 20761 TGGAAGCCAT GCTCAGAAAT GACACCAACG ACCAGTCCTT TAATGACTAC CTTTCCGCCG
 20821 CCAACATGCT ATATCCCATA CCCGCCAACG CCACCAACGT GCCCATCTCC ATCCCATCGC
 20881 GCAAATGGGC AGCATTTCGC GGTTGGGCCT TCACACGCTT GAAGACAAAG GAAACCCCTT
 20941 CCCTGGGATC AGGCTACGAC CCTTAAC TACCTCTGG CTCCATACCA TACCTTGAC
 21001 GAAACTTCTA TCTTAATTCAC ACCTTTAAGA AGGTGGCCAT TACTTTTGAC TCTTCTGTTA
 21061 GCTGGCCGGG CAACGACCGC CTGCTTACTC CCAATGAGTT TGAGATTAAG CGCTCAGTTG
 21121 ACGGGGAGGG CTATAACGTA GCTCAGTGCA ACATGACAAA GGACTGGTTC CTAGTGCAGA
 21181 TGTTGCCAA CTACAATATT GGCTACCAAGG GCTTCTACAT TCCAGAAAGC TACAAAGACC
 21241 GCATGTACTC GTTCTTCAGA AACTTCCAGC CCATGAGCCG GCAAGTGGTG GACGATACTA
 21301 AATAACAAAGA TTATCAGCAG GTTGGAAATTA TCCACCAAGCA TAACAACCTCA GGCTTCGTAG
 21361 GCTACCTCGC TCCCACCATG CGCGAGGGAC AAGCTTACCC CGCTAATGTT CCCTACCCAC
 21421 TAATAGGCAA AACCGCGGTT GATAGTATTA CCCAGAAAAA GTTTCTTGC GACCGCACCC
 21481 TGTGGCGCAT CCCCTCTCC AGTAACTTA TGTCATGGG TGCGCTACCA GACCTGGGCC
 21541 AAAACCTCT CTACGCAAAC TCCGCCAACG CGCTAGACAT GACCTTGAG GTGGATCCCA
 21601 TGGACGAGCC CACCCCTCTT TATGTTTGTT TGAAAGTCTT TGACGTGGTC CGTGTGCACC
 21661 AGCCGCACCG CGCGTCATC GAGACCGTGT ACCTGCGAC GCCCTTCTCG GCGGGCAACG
 21721 CCACAACATA AAGAAGCAAG CAACATCAAC AACAGCTGCC GCCATGGGC CCAGTGAGCA
 21781 GGAACGTAAA GCCATTGTCA AAGATCTGG TTGTGGCCA TATTTTTTGG GCACCTATGA
 21841 CAAGCGCTTC CCAGGCTTTG TTTCCCCACA CAAGCTCGCC TGCGCCATAG TAAACACGGC
 21901 CGGTGGCGAG ACTGGGGCG TACACTGGAT GGCTTGTGCC TGGAACCCGC GCTAAAAAC
 21961 ATGCTACCTC TTGAGCCCT TTGGCTTTTG TGACCAACGT CTCAAGCAGG TTTACCAAGTT
 22021 TGAGTACGAG TCACTCCTGC GCCGTAGCGC CATTGCTCT TCCCCGACC GCTGTATAAC
 22081 GCTGGAAAAG TCCACCCAAA GCGTGCAGGG GCCCAACTCG GCCGCCTGTG GCCTATTCTG
 22141 CTGCATGTTT CTCCACGCC TTGCCAACTG GCCCCAAACT CCCATGGATC ACAACCCAC
 22201 CATGAACCTT ATTACCGGGG TACCAACTC CATGCTTAAC AGTCCCCAGG TACAGCCAC
 22261 CCTGCGCCGC AACCAAGAAC AGCTCTACAG CTTCCCTGGAG CGCCACTCGC CCTACTTCCG
 22321 CAGCCACAGT GCGCAAATTAA GGAGGCCAC TTCTTTTGT CACTGAAAA ACATGTAAAA
 22381 ATAATGTACT AGGAGACACT TTCAATAAG GCAAATGTTT TTATTTGTAC ACTCTCGGGT
 22441 GATTATTTAC CCCCACCTT GCCGTCTGCG CGTTTAAAAA ATCAAAGGGG TTCTGCCCG
 22501 CATCGCTATG CGCCACTGGC AGGGACACGT TGCGATACTG GTGTTTAGTG CTCCACTTAA
 22561 ACTCAGGCAC AACCATCCGC GGCAGCTCGG TGAAGTTTC ACTCCACAGG CTGCCACCA
 22621 TCACCAACGC GTTTAGCAGG TCGGGCGCCG ATATCTTGAA GTCGCAGTTG GGGCCTCCGC

FIG. 7I

22681 CCTGCGCGCG CGAGTTGCAG TACACAGGGT TACAGCACTG GAACACTATC AGCGCCGGGT
 22741 GGTGCACGCT GGCCAGCACG CTCTTGTCGG AGATCAGATC CGCGTCCAGG TCCTCCGCGT
 22801 TGCTCAGGGC GAACGGAGTC AACTTGGTA GCTGCCTTCC CAAAAAGGGT GCATGCCAG
 22861 GCTTTGAGTT GCACTCGCAC CGTAGTGGCA TCAGAAGGTG ACCGTGCCA GTCTGGCGT
 22921 TAGGATACAG CGCCTGCATG AAACCCCTGA TCTGCTTAAA AGCCACCTGA GCCTTGCAG
 22981 CTTCAGAGAA GAACATGCCG CAAGACTTGC CGGAAACTG ATTGGCCGGA CAGGCCCGT
 23041 CATGCACGCA GCACCTTGC CGGGTGTGG AGATCTGCAC CACATTTCGG CCCCACCGT
 23101 TCTTCACGAT CTTGGCCTTG CTAGACTGCT CCTTCAGCGC GCGCTGCCG TTTTCGCTCG
 23161 TCACATCCAT TTCAATCAG TGCTCCTTAT TTATCATAAT GCTCCCGTGT AGACACTTAA
 23221 GCTCGCCTTC GATCTCAGCG CAGCGTGCA GCCACAAACGC GCAGCCCCGTG GGCTCGTGGT
 23281 GCTTGAGGT TACCTCTGCA AACGACTGCA GGTACGCCG CAGGAATCGC CCCATCATCG
 23341 TCACAAAGGT CTTGTTGCTG GTGAAGGTCA GCTGCAACCC GCGGTGCTCC TCGTTTAGCC
 23401 AGGTCTTGCA TACGGCCGCC AGAGCTTCCA CTTGGTCAGG CAGTAGCTG AAGTTGCCT
 23461 TTAGATCGTT ATCCACGTGG TACTTGTCCA TCAACGCCG CGCAGCCTCC ATGCCCTTCT
 23521 CCCACCGAGA CACGATCGGC AGGCTCAGCG GGTITATCAC CGTGCTTCA CTTCCGCTT
 23581 CACTGGACTC TTCCCTTTCC TCTTGCATCC GCATACCCCG CGCCACTGGG TCGTCTTCAT
 23641 TCAGCCGCCG CACCCTGC CGCAGCCTCC TGCCGTGCTT GATTAGCACC GGTGGTTGC
 23701 TGAAACCCAC CATTGTTAGC GCCACATCTT CTCTTCTTC CTCGCTGTCC ACGATCACCT
 23761 CTGGGATGG CGGGCGCTG GGCTTGGAG AGGGCGCTT CTTTTCTTT TTGGACGCAA
 23821 TGGCCAAATC CGCCGTGAG GTCGATGCC GCGGGCTGGG TGTGCGCCGC ACCAGCGCAT
 23881 CTTGTACGA GTCTTCTTCG TCCTCGGACT CGAGACGCCG CCTCAGCCGC TTTTTGGGG
 23941 GCGCGGGGG AGGCGGGGGC GACGGCGACG GGGACGAGAC GTCCCTCCATG GTTGGTGGAC
 24001 GTCGCCGCC ACCCGCTCCG CGCTCGGGG TGTTTGC CGCTCCTCT TCCCGACTGG
 24061 CCATTTCTT CTCCCTATAGG CAGAAAAGA TCATGGAGTC AGTCGAGAAG GAGGACAGCC
 24121 TAACCCCCCC CTTTGAGTTC GCCACCACCG CCTCCACCGA TGCCGCCAAC GCGCTTACCA
 24181 CCTTCCCCGT CGAGGCACCC CCGCTTGAGG AGGAGGAAGT GATTATCGAG CAGGACCCAG
 24241 GTTTGTAAG CGAAGACGAC GAAGATCGCT CAGTACCAAC AGAGGATAAA AAGCAAGACC
 24301 AGGACGACGC AGAGGCAAAC GAGGAACAAG TCGGGGGGG GGACCAAAGG CATGGCACT
 24361 ACCTAGATGT GGGAGACGAC GTGCTGTTGA AGCATCTGCA GCGCCAGTGC GCCATTATCT
 24421 GCGACCCGTT GCAAGAGCGC AGCGATGTGC CCCTCGCCAT AGCGGATGTC AGCCTTGCCT
 24481 ACGAACGCCA CCTGTTCTCA CGCGCGTAC CCCCCAAACG CCAAGAAAAC GGCACATGCG
 24541 AGCCCAACCC GCGCCTAAC TTCTACCCCG TATTGCGCT GCGAGGGTG CTTGCCACCT
 24601 ATCACATCTT TTTCCAAAC TGCAAGATAC CCCTATCCTG CCGTGCCAAAC CGCAGCCGAG
 24661 CGGACAAGCA GCTGGCCTTG CGGCAGGGCG CTGTCATACC TGATATGCC TCGCTCGACG
 24721 AAGTGCCAAA AATCTTGAG GGTCTTGGAC GCGACGAGAA GCGCGCGGCA AACGCTCTGC
 24781 AACAAAGAAA CAGCGAAAAT GAAAGTCACT GTGGAGTGCT GGTGGAACCTT GAGGGTGACA
 24841 ACGCGCCCT AGCCGTGCTG AAACGCAGCA TCGAGGTCAC CCACCTTGCC TACCCGGCAC
 24901 TTAACCTACC CCCCCAGGTT ATGAGCACAG TCATGAGCGA GCTGATCGTG CGCCGTGCAC
 24961 GACCCCTGGA GAGGGATGCA AACTTGCAAG AACAAACCGA GGAGGGCCTA CCCGAGTTG
 25021 GCGATGAGCA GCTGGCGCGC TGGCTTGAGA CGCCGAGGCC TGCCGACTTG GAGGAGCGAC
 25081 GCAAGCTAAT GATGGCCGCA GTGCTTGTAA CGTGGAGCT TGAGTGCATG CAGCGGTTCT
 25141 TTGCTGACCC GGAGATGCAG CGCAAGCTAG AGGAAACGTT GCACTACACC TTTGCCAGG

52/92

25201 GCTACGTGCG CCAGGCCTGC AAAATTTCCA ACGTGGAGCT CTGCAACCTG GTCTCCTACC
 25261 TTGGAATTTT GCACGAAAAC CGCCTTGGC AAAACGTGCT TCATTCCACG CTCAAGGGCG
 25321 AGGCAGGCCG CGACTACGTC CGCGACTGCG TTTACTTATT TCTGTGCTAC ACCTGGCAA
 25381 CGGCCATGGG CGTGTGGCAG CAGTGCCTGG AGGAGCGCAA CCTGAAGGAG CTGCAGAAC
 25441 TGCTAAAGCA AAACTTGAAG GACCTATGGA CGGCCTCAA CGAGCGCTCC GTGGCCGCG
 25501 ACCTGGCGGA CATTATCTTC CCCGAACGCC TGCTAAAAC CCTGCAACAG GGTCTGCCAG
 25561 ACTTCACCAAG TCAAAGCATG TTGAAAAC TTAGGAACCTT TATCCTAGAG CGTTCAGGAA
 25621 TTCTGCCCAG CACCTGCTGT GCGCTTCTA GCGACTTGT GCCCATTAAG TACCGTGAAT
 25681 GCCCTCCGCC GCTTGGGGT CACTGCTACC TTCTGAGCT AGCCAACTAC CTTGCTTAC
 25741 ACTCCGACAT CATGGAAGAC GTGAGCGGTG ACGGCCTACT GGAGTGTAC TGCGCTGCA
 25801 ACCTATGCAC CCCGCACCGC TCCCTGGTCT GCAATTACA ACTGCTTAGC GAAAGTCAAA
 25861 TTATCGGTAC CTTTGAGCTG CAGGGTCCCT CGCCTGACGA AAAGTCCGCG GCTCCGGGT
 25921 TGAAACTCAC TCCGGGGCTG TGGACGTCCG CTTACCTTCG CAAATTTGTA CCTGAGGACT
 25981 ACCACGCCCA CGAGATTAGG TTCTACGAAG ACCAATCCCG CCCGCCAAAT GCGGAGCTTA
 26041 CCGCCTGCGT CATTACCCAG GGCCACATCC TTGGCCAATT GCAAGCCATT AACAAAGCCC
 26101 GCCAAGAGTT TCTGCTACGA AAGGGACGGG GGTTTACTT GGACCCCCAG TCCGGCGAGG
 26161 AGCTCAACCC AATCCCCCG CGCCGCAGC CCTATCAGCA GCCGCGGGCC CTTGCTTCCC
 26221 AGGATGGCAC CCAAAAAGAA GCTGCAGCTG CGGCCGCCGC CACCCACGGA CGAGGAGGAA
 26281 TACTGGGACA GTCAGGCAGA GGAGGTTTG GACGAGGAGG AGGAGATGAT GGAAGACTGG
 26341 GACAGCCTAG ACGAGGAAGC TTCCGAGGCC GAAGAGGTGT CAGACGAAAC ACCGTCACCC
 26401 TCGGTCGCAT TCCCCTCGCC GGCGCCCCAG AAATCGCAA CCGTTCCAG CATTGCTACA
 26461 ACCTCCGCTC CTCAGGCAGC GCCGGCACTG CCCGTTCGCC GACCCAACCG TAGATGGGAC
 26521 ACCACTGGAA CCAGGGCCGG TAAGTCTAAG CAGCCGCCGC CGTTAGCCA AGAGCAACAA
 26581 CAGCGCCAAG GCTACCGCTC GTGGCGCGTG CACAAGAACG CCATAGTTGC TTGCTTGAA
 26641 GACTGTGGGG GCAACATCTC CTTGGCCCGC CGTTTCTTC TCTACCATCA CGCGTGGCC
 26701 TTCCCCCGTA ACATCCTGCA TTACTACCGT CATCTCTACA GCCCCTACTG CACCGGCGGC
 26761 AGCGGAGCA ACAGCAGCGG CCACGCAGAA GCAAAGGCAG CGGGATAGCA AGACTCTGAC
 26821 AAAGCCAAG AAATCCACAG CGGCAGCAGC AGCAGGAGGA GGAGCACTGC GTCTGGCGCC
 26881 CAACGAACCC GTATCGACCC GCGAGCTTAG AAACAGGATT TTTCCCACTC TGTATGCTAT
 26941 ATTTCAACAG AGCAGGGGCC AAGAACAAAG GCTGAAAATA AAAAACAGGT CTCTGCGCTC
 27001 CCTCACCCGC AGCTGCCTGT ATCACAAAG CGAACATCAG CTTGGCGCA CGCTGGAAGA
 27061 CGCGGAGGCT CTCTTCAGCA AATACTGCAG CGTACTCTT AAGGACTAGT TTCGGCCCT
 27121 TTCTCAAATT TAAGCGCGAA AACTACGTCA TCTCCAGCGG CCACACCCCG CGCCAGCACC
 27181 TGTCGTACAGC GCCATTATGA GCAAGGAAAT TCCCACGCC TACATGTGGA GTTACCAAGCC
 27241 ACAAAATGGGA CTTGGGGCTG GAGCTGCCA AGACTACTCA ACCCGAATAA ACTACATGAG
 27301 CGCGGGACCC CACATGATAT CCCGGGTCAA CGGAATCCGC GCCCACCGAA ACCGAATTCT
 27361 CCTCGAACAG CGGGCTATTA CCACCAACACC TCGTAATAAC CTTAACCCCC GTAGTTGGCC
 27421 CGCTGCCCTG GTGTACCGAG AAAGTCCCGC TCCCACCACT GTGGTACTTC CCAGAGACGC
 27481 CCAGGGCGAA GTTCAGATGA CTAACCTCAGG GGCGCAGCTT GCGGGCGGGCT TTGTCACAG
 27541 GGTGCGGTGCG CCCGGGCAGG GTATAACTCA CCTGAAAATC AGAGGGCGAG GTATTCAAGCT
 27601 CAACGACGAG TCGGTGAGCT CCTCTCTTGG TCTCCGTCCG GACGGGACAT TTCAGATCGG
 27661 CGGCCTGGC CGCTCTTCAT TTACGCCCG TCAGGCGATC CTAACCTCTGC AGACCTCGTC

FIG. 7K

27721 CTCGGAGCCG CGCTCCGGAG GCATTGGAAC TCTACAATT ATTGAGGAGT TCGTGCCTTC
 27781 GGTTTACTTC AACCCCTTTT CTGGACCTCC CGGCCACTAC CGGGACCAGT TTATCCCCAA
 27841 CTTTGACGCG GTAAAAGACT CGGCGGACGG CTACGACTGA ATGACCAGTG GAGAGGCAGA
 27901 GCAACTGCGC CTGACACACC TCGACCACTG CCCGCCAAC AAGTGCTTTG CCCGCCGCTC
 27961 CGGTGAGTTT TGTTACTTTG AATTGCCGA AGAGCATATC GAGGGCCCGG CGCACGGCGT
 28021 CGGGCTCACC ACCCAGGTAG AGCTTACACG TAGCCTGATT CGGGAGTTTA CCAAGCGCCC
 28081 CCTGCTAGTG GAGCGGGAGC GGGGTCCCTG TGTCTGACC GTGGTTGCA ACTGTCTAA
 28141 CCCTGGATTAA CATCAAGATC TTTGTTGTCA TCTCTGTGCT GAGTATAATA AATACAGAAA
 28201 TTAGAATCTA CTGGGGCTCC TGTGCCATC CTGTGAACGC CACCCTTTT ACCCACCAA
 28261 AGCAGACCAA AGCAACCTC ACCTCCGGTT TGACACAAGCG GGCAATAAG TACCTTACCT
 28321 GGTACTTTAA CGGCTCTTCA TTTGTAATT ACAACAGTT CCAGCGAGAC GAAGTAAGTT
 28381 TGCCACACAA CCTTCTCGGC TTCAACTACA CCGTCAAGAA AAACACCACC ACCACCCCTCC
 28441 TCACCTGCCG GGAACGTACG AGTGCCTCAC CGGTGCTGC GCCCCACACCT ACAGCCTGAG
 28501 CGTAACCAGA CATTACTCCC ATTTCCCAA AACAGGAGGT GAGCTCAACT CCCGAACTC
 28561 AGGTCAAAAA AGCATTTCG GGGGTGCTGG GATTTTTAA TTAAGTATAT GAGCAATTCA
 28621 AGTAACCTCTA CAAGCTTGTC TAATTTTCT GGAATTGGGG TCGGGGTTAT CCTTACTCTT
 28681 GTAATTCTGT TTATTCTTAT ACTAGCACTT CTGTCCTTA GGGTTGCCGC CTGCTGCACG
 28741 CACGTTGTA CCTATTGTCA GCTTTTAA CGCTGGGGC GACATCCAAG ATGAGGTACA
 28801 TGATTTTAGG CTGCTCGCC CTTGCGGCAG TCTGCAGCGC TGCCAAAAAG GTTGAGTTA
 28861 AGGAACCAGC TTGCAATGTT ACATTTAAAT CAGAAGCTAA TGAATGCACT ACTCTTATAA
 28921 AATGCCAAC AGAACATGAA AAGCTTATTA TTGCCACAA AGACAAAATT GGCAAGTATG
 28981 CTGTATATGC TATTTGGCAG CCAGGTGACA CTAACGACTA TAATGTACA GTCTCCAAG
 29041 GTGAAAATCG TAAAACTTT ATGTATAAT TTCCATTGTT TGAAATGTGC GATATTACCA
 29101 TGTACATGAG CAAACAGTAC AAGTTGTGGC CCCCACAAAA GTGTTAGAG AACACTGGCA
 29161 CCTTTGTTT CACCGCTCTG CTTATTACAG CGCTGCTTT GGTATGTACC TTACTTTATC
 29221 TCAAATACAA AAGCAGACGC AGTTTATTG ATGAAAAGAA AATGCCTTGA TTTTCCGCTT
 29281 GCTTGTATTTC CCTGGACAA TTTACTCTAT GTGGGATATG CGCCAGGCAG GAAAGATTAT
 29341 ACCCACAACC TTCAAATCAA ACTTTCTGG ACGTTAGCGC CTGACTTCTG CCAGCGCCTG
 29401 CACTGCAAAT TTGATCAAAC CCAGCTTCAG CTTGCCCTGCT CCAGAGATGA CCGGCTCAAC
 29461 CATCGCGCC ACAACGGACT ATCGAACAC CACTGCTACC GGACTAAAAT CTGCCCTAAA
 29521 TTTACCCCAA GTTCATGCCT TTGTCAATGA CTGGCGAGC TTGGGCATGT GGTGGTTTC
 29581 CATAGCGCTT ATGTTTGTGTT GCCTTATTAT TATGTGGCTT ATTTGTTGCC TAAAGCGCAG
 29641 ACAGCGCAGA CCCCCCATCT ATAGGCCTAT CATTGTGCTC AACCCACACA ATGAAAAAAAT
 29701 TCATAGATTG GACGGTCTCA AACCATGTTT CTTCTTTTA CAGTATGATT AAATGAGACA
 29761 TGATTCTCG AGTCCTTATA TTATTGACCC TTGTTGCGCT TTTCTGTGCG TGCTCTACAT
 29821 TGGCTGCCGGT CGCTCACATC GAAGTAGATT GCATCCCACC TTTCACAGTT TACCTGCTTT
 29881 ACGGATTGT CACCCCTTATC CTCATCTGCA GCCTCGTCAC TGTAGTCATC GCCTTCATTC
 29941 AGTCATTGA CTGGATTGT GTGCGCATTG CGTACCTTAG GCACCACCG CAATACAGAG
 30001 ACAGGACTAT AGCTGATCTT CTCAGAATTC TTTAATTATG AAACGGATTG TCACTTTGT
 30061 TTTGCTGATT TTCTGCGCCC TACCTGTGCT TTGCTCCCAA ACCTCAGCGC CTCCCAAAAG
 30121 ACATATTCC TGCAGATTCA CTCAAATATG GAACATTCCC AGCTGCTACA ACAAACAGAG
 30181 CGATTGTCA GAAGCCTGGT TATAGCCAT CATCTGTGTC ATGGTTTTT GCAGTACCAT

30241 TTTGCCCTA GCCATATAACC CATAACCTTGA CATTGGTTGG AATGCCATAG ATGCCATGAA
 30301 CCACCCCTACT TTCCCAGCGC CCAATGTCAT ACCACTGCAA CAGGTTATTG CCCCAATCAA
 30361 TCAGCCTCGC CCCCCCTTCTC CCACCCCCAC TGAGATTAGC TACTTTAATT TGACAGGTGG
 30421 AGATGACTGA ATCTCTAGAT CTAGAATTGG ATGGAATTAA CACCGAACAG CGCCTACTAG
 30481 AAAGGCAGCAA GGCAGCGTCC GAGCGAGAAC GCCTAAAACA AGAAGTTGAA GACATGGTTA
 30541 ACCTGCACCA GTGTAAAAGA GGTATCTTT GTGTGGTCAA GCAGGCCAAA CTTACCTACG
 30601 AAAAAACAC TACCGGCAAC CGCCTTAGCT ACAAGCTACC CACCCAGCGC CAAAAACTGG
 30661 TGCTTATGGT GGGAGAAAAA CCTATCACCG TCACCCAGCA CTCGGCAGAA ACAGAAGGCT
 30721 GCCTGCACCT CCCCTATCAG GGTCCAGAGG ACCTCTGCAC TCTTATTAAA ACCATGTGTG
 30781 GCATTAGAGA TCTTATTCCA TTCAACTAAC AATAAACACA CAATAAAATTA CTTACTTAAA
 30841 ATCAGTCAGC AAATCTTGT CCAGCTTATT CAGCATCACC TCCTTTCCCT CCTCCCAACT
 30901 CTGGTATTTC AGCAGCCTTT TAGCTGCCAA CTTTCTCCAA AGTCTAAATG GGATGTCAAA
 30961 TTCCTCATGT TCTTGTCCTC CGCGACCCAC TATCTTCATA TTGTTGCAGA TGAAACGCGC
 31021 CAGACCGTCT GAAGACACCT TCAACCCGT GTACCCATAT GACACGGAA CGGGCCCTCC
 31081 AACTGTGCCT TTCCCTTACCC CTCCCTTGT GTGCCAAAT GGGTTCCAAG AAAGTCCCCC
 31141 CGGAGTGCCT TCTTGCGTC TTTCAGAACC TTTGGTTACC TCACACGGCA TGCTTGCGCT
 31201 AAAAAATGGGC AGCGGCCTGT CCCTGGATCA GGCAGGCAAC CTTACATCAA ATACAATCAC
 31261 TGTTCCTCAA CCGCTAAAAA AAACAAAGTC CAATATAACT TTGGAAACAT CGCGCCCTCC
 31321 TACAGTCAGC TCAGGGGCCCC TAACCATGGC CACAACCTCG CCTTTGGTGG TCTCTGACAA
 31381 CACTCTTACC ATGCAATCAC AAGCACCGCT AACCGTGCAA GACTCAAAAC TTAGCATTGC
 31441 TACCAAAGAG CCACCTTACAG TGTTAGATGG AAAACTGGCC CTGCAGACAT CAGCCCCCT
 31501 CTCTGCCACT GATAACAACG CCCTCACTAT CACTGCCCTCA CCTCCTCTTA CTACTGCAAA
 31561 TGGTAGTCTG GCTGTTACCA TGGAAAACCC ACTTTACAAC AACAATGGAA AACTTGGGCT
 31621 CAAAATTGGC GGTCTTTGC AAGTGGCCAC CGACTCACAT GCACTAACAC TAGGTACTGG
 31681 TCAGGGGGTT GCAGTTCATA ACAATTGCT ACATACAAAA GTTACAGGGC CAATAGGGTT
 31741 TGATACATCT GGCAACATGG AACTTAAAAC TGGAGATGGC CTCTATGTGG ATAGGCCGG
 31801 TCCTAACCAA AAACATACATA TTAATCTAAA TACCACAAAA GGCTTGCTT TTGACAACAC
 31861 CGCAATAACA ATTAACGCTG GAAAAGGGTT GGAATTGAA ACAGACTCCT CAAACGGAAA
 31921 TCCCATAAAA ACAAAAATTG GATCAGGCAT ACAATATAAT ACCAATGGAG CTATGGTTGC
 31981 AAAACTTGGG ACAGGCCCTCA GTTTGACAG CTCCGGAGCC ATAACAATGG GCAGCATAAA
 32041 CAATGACAGA CTTACTCTTT GGACAACACC AGACCCATCC CCAAATTGCA GAATTGCTTC
 32101 AGATAAAAGAC TGCAAGCTAA CTCTGGCGCT AACAAAATGT GGCAGTCAAA TTTTGGGCAC
 32161 TGTTTCAGCT TTGGCAGTAT CAGGTAATAT GGCTTCCATC AATGGAACTC TAAGCAGTGT
 32221 AAACTTGGTT CTTAGATTG ATGACAACGG AGTGCCTTATG TCAAATTCAT CACTGGACAA
 32281 ACAGTATTGG AACTTTAGAA ACGGGGACTC CACTAACGGT CAACCATAAC CTTATGCTGT
 32341 TGGGTTTATG CCAAACCTAA AAGCTTACCC AAAACTCAA AGTAAAACGT CAAAAGTAA
 32401 TATTGTTAGC CAGGTGTATC TTAATGGTGA CAAGTCTAAA CCATTGCATT TTACTATTAC
 32461 GCTAAATGGA ACAGATGAAA CCAACCAAGT AAGCAAATAC TCAATATCAT TCAGTGGTC
 32521 CTGGAACAGT GGACAATACA CTAATGACAA ATTIGCCACC AATTCCCTATA CCTTCTCCTA
 32581 CATTGCCAG GAATAAAGAA TCGTGAACCT GTTGCATGTT ATGTTTCAAC GTGTTTATT
 32641 TTCAATTGCA GAAAATTTCAGTCA AGTCATTGTT CATTCACTAG TATAGCCCCA CCACCCACATA
 32701 GCTTATACTA ATCACCGTAC CTTAATCAAAC CTCACAGAAC CCTAGTATTG AACCTGCCAC

32761 CTCCCTCCCC ACACACAGAG TACACAGTCC TTTCTCCCCG GCTGGCCTTA AACAGCATCA
32821 TATCATGGGT AACAGACATA TTCTTAGGTG TTATATTCCA CACGGCTCCTC TGTGAGGCCA
32881 AACGCTCATC AGTGATGTTA ATAAACTCCC CGGGCAGCTC GCTTAAGTTC ATGTCGCTGT
32941 CCAGCTGCTG AGCCACAGGC TGCTGTCCAA CTTGCGGTTG CTCAACGGGC GGCGAAGGAG
33001 AAGTCCACGC CTACATGGGG GTAGAGTCAT AATCGTGCAT CAGGATAGGG CGGTGGTGCT
33061 GCAGCAGCGC GCGAATAAAC TGCTGCCGCC GCGCCTCCGT CCTGCAGGAA TACAACATGG
33121 CAGTGGTCTC CTCAGCGATG ATTGCAACCG CCCGCAGCAT AAGGCGCTT GTCCCTCCGGG
33181 CACAGCAGCG CACCCGTGATC TCACCTTAAGT CAGCACAGTA ACTGCAGCAC AGTACCACAA
33241 TATTGTTAA AATCCCACAG TGCAAGCGC TGATCCAAA GCTCATGGCG GGGACCACAG
33301 AACCCACGTG GCCATCATAAC CACAACCGCA GGTAGATTAA GTGGCGACCC CTCATAAACAA
33361 CGCTGGACAT AAACATTACC TCTTTGGCA TGTTGTAATT CACCACCTCC CGGTACCATCA
33421 TAAACCTCTG ATTAAACATG GCGCCATCCA CCACCACCT AAACCAAGCTG GCCAAAACCT
33481 GCCCCGCCGGC TATGCACTGC AGGGAACCGG GACTGGAACA ATGACAGTGG AGAGCCCAGG
33541 ACTCGTAACC ATGGATCATC ATGCTCGTCA TGATATCAAT GTTGGCACAA CACAGGCACA
33601 CGTGCACTACA CTCCTCGAGG ATTACAAGCT CCTCCCGCGT CAGAACCATATA TCCCAGGGAA
33661 CAACCCATTC CTGAATCAGC GTAAATCCC CACTGCAGGG AAGACCTCGC ACGTAACCTCA
33721 CGTTGTGCA TGTCAAAGTG TTACATTGG GCAGCAGCGG ATGATCCTCC AGTATGGTAG
33781 CGCGTGTCTC TGTCTCAAAA GGAGGTAGGC GATCCCTACT GTACGGAGTG CGCCGAGACA
33841 ACCGAGATCG TGGTGGTCGT AGTGTCACTGC CAAATGGAAC GCCGGACGTA GTCATATTTC
33901 CTGAAGCAAA ACCAGGTGCG GGCAGTACAA ACAGATCTGC GTCTCCGGTC TCGTCGCTTA
33961 GCTCGCTCTG TGTAGTAGTT GTAGTATATC CACTCTCTCA AAGCATCCAG GCGCCCCCTG
34021 GCTTCGGGTT CTATGTAAAC TCCTTCATGC GCGCCTGCC TGATAACATC CACCACCGCA
34081 GAATAAGCCA CACCCAGCCA ACCTACACAT TCGTTCTGCG AGTCACACAC GGGAGGAGCG
34141 GGAAGAGCTG GAAGAACCAT GTTTTTTTTT TTTATTCCAA AAGATTATCC AAAACCTCAA
34201 AATGAAGATC TATTAAGTGA ACAGCGCTCCC CTCCGGTGGC GTGGTCAAAC TCTACAGCCA
34261 AAGAACAGAT AATGGCATTG GTAAGATGTT GCACAATGGC TTCCAAAAGG CAAACTGCC
34321 TCACGTCAA GTGGACGTAA AGGCTAAACC CTTCAAGGGTG AATCTCCTCT ATAAACATTC
34381 CAGCACCTTC AACCATGCCA AAATAATTTC CATCTGCCA CCTTATCAAT ATGTCCTAA
34441 GCAAATCCCG AATATTAAGT CGGGCCATTG TAAAAATCTG CTCCAGAGCG CCCTCCACCT
34501 TCACGTCAA GCAGCGAATC ATGATTGCAA AAATTCAAGGT TCCTCACAGA CCTGTATAAG
34561 ATTCAAAAGC GGAACATTAA CAAAAATACC GCGATCCCGT AGGTCCCTTC GCAGGGCCAG
34621 CTGAACATAA TCGTGCAGGT CTGCACGGAC CAGCGCGGCC ACTTCCCCGC CAGGAACCAT
34681 GACAAAAGAA CCCACACTGA TTATGACACG CATACTCGGA GCTATGCTAA CCAGCGTAGC
34741 CCCGATGTAA GCTTGTGCA TGGGCAGCGA TATAAAATGC AAGGTACTGC TCAAAAAATC
34801 AGGAAAGCC TCGCGAAAAA AAGCAAGCAC ATCGTAGTCA TGCTCATGCA GATAAAGGCA
34861 GGTAAGTTCC GGAACCACCA CAGAAAAAGA CACCAATTTC CTCTCAAAACA TGTCTGCGGG
34921 TTCCTGCATA AACACAAAAT AAAATAACAA AAAAAAAA ACATTTAAC ATTAGAAGCC
34981 TGTNTTACAA CAGGAAAAAC AACCCCTATA AGCATAAGAC GGACTACGGC CATGCCGGCG
35041 TGACCGTAA AAAACTGGTC ACCGTGATTA AAAAGCACCA CCGACAGTTC CTCGGTCATG
35101 TCCGGAGTCA TAATGTAAAGA CTCGGTAAAC ACATCAGGTT GGTTAACATC GGTCAGTGCT
35161 AAAAGCGAC CGAAATAGCC CGGGGGAAATA CATAACCGCA GGCGTAGAGA CAACATTACA
35221 GCCCCCATAG GAGGTATAAC AAAATTAAATA GGAGAGAAAA ACACATAAAC ACCTGAAAAA

56/92

35281 CCCTCCGTGCC TAGGCAAAAT AGCACCCCTCC CGCTCCAGAA CAACATACAG CGCTTCCACA
35341 GCGGCAGCCA TAACAGTCAG CCTTACCACTT AAAAAAACCT ATTAAAAAAAC ACCACTCGAC
35401 ACGGCACCCAG CTCAAATCAGT CACAGTGTA AAAGGGCCAA GTACAGAGCG AGTATATATA
35461 GGACTAAAAA ATGACGTAAC GGTTAAAGTC CACAAAAACC ACCCAGAAAA CCGCACCGGA
35521 ACCTACGCC AGAAACGAAA GCCAAAAAAC CCACAACTTC CTCAAATCTT CACTCCGTT
35581 TTCCCCACGAT ACGTCACTTC CCATTTAAA AAAAAACTAC AATTCCCAAT ACATGCAAGT
35641 TACTCCGCC TAAAACCTAC GTCACCCGCC CCGTCCCAC GCCCCGCC ACGTACAAA
35701 CTCCACCCCC TCATTATCAT ATTGGCTTCA ATCCAAAATA AGGTATATTA TTGATGATG

FIG. 7O

1 CATCATCAAT AATATAACCTT ATTTTGAGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT
 61 TTGTGACGTG GCGCGGGGGC TGGGAACGGG GCGGGTGACG TAGTAGTGTG GCAGAAAGTGT
 121 GATGTTGCAA GTGTGGCGGA ACACATGTAA GCGACGGATG TGGCAAAAGT GACGTTTTG
 181 GTGTGCCCG GTGTACACAG GAAGTGACAA TTTTCGCGCG GTTTTAGGCG GATGTTGTAG
 241 TAAATTGGG CGTAACCGAG TAAGATTGG CCATTTTCGC GGGAAAACGT AATAAGAGGA
 301 AGTGAAATCT GAATAATTTT GTGTTACTCA TAGCGCGTAA TATTGCTA GGGCCGCGGG
 361 GACTTGACC GTTACGTGG AGACTCGCCC AGGTGTTTT CTCAGGTGTT TTCCGCGTTC
 421 CGGGTCAAAG TTGGCGTTT ATTATTATAG TCAGCTGACG TGTAGTGTAT TTATACCCGG
 481 TGAGTCTCCTC AAGAGGCCAC TCTTGAGTGC CAGCGAGTAG AGTTTCTCC TCCGAGCCGC
 541 TCCGACACCG GGACTGAAAA TGAGACATAT TATCTGCCAC GGAGGTGTTA TTACCGAAGA
 601 AATGGCCGCC AGTCTTTGG ACCAGCTGAT CGAAGAGGTA CTGGCTGATA ATCTTCCACC
 661 TCCTAGCCAT TTGAAACCAC CTACCCCTCA CGAACTGTAT GATTTAGACG TGACGGCCCC
 721 CGAAGATCCC AACGAGGGAGG CGGTTTCGCA GATTTTCCC GACTCTGAA TGTTGGCGGT
 781 GCAGGAAGGG ATTGACTTAC TCACTTTCC GCGGGCGCCC GTTCTCCGG AGCCGCCTCA
 841 CCTTTCCCGG CAGCCCGAGC AGCCGGAGCA GAGAGCCTG GGTCCGGTTT CTATGCCAAA
 901 CCTTGACCG GAGGTGATCG ATCTTACCTG CCACGAGGCT GGCTTCCAC CCAGTGACGA
 961 CGAGGATGAA GAGGGTGAGG AGTTTGTGTT AGATTATG TGAGCACCCCG GGCACGGTTG
 1021 CAGGTCTTGT CATTATCACC GGAGGAATAC GGGGGACCA GATATTATGT GTTGCCTTTG
 1081 CTATATGAGG ACCTGTGGCA TGTTTGTCTA CAGTAAGTGA AAATTATCGG CAGTGGGTGA
 1141 TAGAGTGGTG GTTGGTGTGTT GGTAATTTTT TTGTTAATT TTACAGTTT GTGGTTAAA
 1201 GAATTTGTA TTGTTGATTTT TTGTTAAAGGT CCTGTGTCTG AACCTGAGCC TGACCCGAG
 1261 CCAGAACCCG AGCCTGCAAG ACCTACCCCG CGTCTAAAAA TGGCGCCTGC TATCTGAGA
 1321 CGCCCGACAT CACCTGTGTC TAGAGAAATGC AATAGTAGTA CGGATAGCTG TGACTCCGGT
 1381 CCTTCTAACCA CACCTCCCTGA GATACACCCCG GTGGTCCCGC TGTGCCCAT TAAACCAAGTT
 1441 GCCGTGAGAG TTGGTGGGGC TCGCCAGGCT GTGGAATGTA TCGAGGACTT GCTTAACGAG
 1501 CCTGGGCAAC CTTGGACTT GAGCTGTAAGG CGCCCCAGGC CATAAGGTGT AAACCTGTGA
 1561 TTGCGTGTGTG GTTAAACGCC TTGTTGCT GAATGAGTTG ATGTAAGTTT AATAAGGGT
 1621 GAGATAATGTT TTAACCTGCA TGGCGTGTAA ATGGGGCGG GGCTTAAAGG GTATATAATG
 1681 CGCCGTGGGC TAATCTTGGT TACATCTGAC CTCATGGAGG CTGGGAGTG TTTGGAAGAT
 1741 TTTCTGCTG TGCGTAACCTT GCTGGAACAG AGCTCTAACAA GTACCTCTTG GTTTGGAGG
 1801 TTTCTGTGG GCTCATCCCA GGCAAAAGTTA GTCTGCAGAA TTAAGGAGGA TTACAAGTGG
 1861 GAATTTGAAAG AGCTTTGAA ATCCCTGTGGT GAGCTGTTTG ATTCTTGAA TCTGGTCAC
 1921 CAGGCCCTTT TCCAAGAGAA GGTCTCATCAAG ACTTTGGATT TTTCCACACC GGGCGCGCT
 1981 GCGGCTGCTG TTGCTTTTTT GAGTTTATA AAGGATAAAAT GGAGCGAAGA AACCCATCTG
 2041 AGCGGGGGGT ACCTGCTGGA TTTCTGGCC ATGCATCTGT GGAGAGCGGT TGTGAGACAC
 2101 AAGAACCGCC TGCTACTGTT GTCTTCCGTC CGCCCCGGCGA TAATACCGAC GGAGGAGCAG
 2161 CAGCAGCAGC AGGAGGAAGC CAGGGGGCGG CGGCAGGAGC AGAGCCCATG GAACCCGAGA
 2221 GCCGGCCTGG ACCCTCGGGA ATGAATGTTG TACAGGTGGC TGAACGTAT CCAGAACTGA
 2281 GACGCATTTC GACAATTACA GAGGATGGC AGGGGCTAAA GGGGGTAAAG AGGGAGCGGG
 2341 GGGCTGTGA GGCCTACAGAG GAGGCTAGGA ATCTAGCTTT TAGCTTAATG ACCAGACACC
 2401 GTCCTGAGTG TATTACTTTT CAACAGATCA AGGATAATTG CGCTAATGAG CTTGATCTGC
 2461 TGGCGAGAA GTATTCCATA GAGCAGCTGA CCACTTACTG GCTGCAGCCA GGGGATGATT
 2521 TTGAGGAGGC TATTAGGTA TATGCAAAGG TGGCACTTAG GCCAGATTGC AAGTACAAGA
 2581 TCAGCAAATCT TGTAATATC AGGAATTGTT GCTACATTTT TGGGAACGGG GCGGAGGTGG
 2641 AGATAGATAC GGAGGATAGG GTGGCCTTTA GATGTAGCAT GATAAATATG TGGCCGGGGG
 2701 TGCTTGGCAT GGACGGGGTG GTTATTATGA ATGTAAGGTT TACTGGCCCC AATTTAGCG
 2761 GTACGGTTTT CCTGGCCAAT ACCAACCTTA TCCTACACGG TGTAAGCTTC TATGGTTTA
 2821 ACAATACCTG TGTTGGAGCC TGGACCGATG TAAGGGTTCG GGGCTGTGCC TTTTACTGCT
 2881 GCTGGAAAGGG GGTGGTGTGT CGCCCCAAAAA GCAGGGCTTC AATTAAGAAA TGCCTCTTTG
 2941 AAAGGTGTAC CTTGGGTATC CTGTCTGAGG GTAACCTCCAG GGTGCGCCAC AATGTGGCCT
 3001 CCGACTGTGG TTGCTTCATG CTAGTAAAAA GCGTGGCTGT GATTAAGCAT AACATGGTAT
 3061 GTGGCAACTG CGAGGACAGG GCCTCTCAGA TGCTGACCTG CTCGGACGGC AACTGTCACC
 3121 TGCTGAAGAC CATTACAGTA GCCAGCCACT CTCGCAAGGC CTGGCCAGTG TTTGAGCATA
 3181 ACATACTGAC CCGCTGTTCC TTGCAATTGG GTAAACAGGAG GGGGGTGTTC CTACCTTACC
 3241 AATGCAATTG GAGTCACACT AAGATATTGC TTGAGCCCGA GAGCATGTCC AAGGTGAACC

3301 TGAACGGGT GTTGACATG ACCATGAAGA TCTGGAAGGT GCTGAGGTAC GATGAGACCC
 3361 GCACCAGGTG CAGACCCCTGC GAGTGTGGCG GTAAACATAT TAGGAACCAG CCTGTGATGC
 3421 TGGATGTGAC CGAGGAGCTG AGGCCCGATC ACTTGGTGCT GCCCTGCACC CGCGCTGAGT
 3481 TTGGCTCTAG CGATGAAGAT ACAGATTGAG GTACTGAAAT GTGTGGCGT GGCTTAAGGG
 3541 TGGGAAAGAA TATATAAGGT GGGGGTCTTA TGTAGTTTG TATCTGTTT GCAGCAGCCG
 3601 CCGCCGCCAT GAGCACCAAC TCGTTTGATG GAAGCATTGT GAGCTCATAT TTGACAACGC
 3661 GCATGCCCGG ATGGGGCGGG GTGCGTCAGA ATGTGATGGG CTCCAGCATT GATGGTCGCC
 3721 CCGTCCTGCC CGCAAACCTCT ACTACCTGA CCTACGAGAC CGTGTCTGGA ACGCCGTTGG
 3781 AGACTGCAGC CTCCGCCGCC GCTTCAGCCG CTGAGCCAC CGCCCGCGGG ATTGTGACTG
 3841 ACTTTGCTT CCTGAGCCCG CTTGCAAGCA GTGAGCCTTC CGGTTCATCC GCCCCGCGATG
 3901 ACAAGTTGAC GGTCTTTTG GCACAATTGG ATTCTTGAC CCGGGAACTT AATGTCGTTT
 3961 CTCAGCAGCT GTTGGATCTG CGCCAGCAGG TTTCTGCCCT GAAGGCTTCC TCCCCCTCCA
 4021 ATGCGGTTTA AAACATAAAAT AAAAACCCAG ACTCTGTTTG GATTGGATC AAGCAAGTGT
 4081 CTTGCTGTCT TTATTTAGGG GTTTTGCAGCG CGCGGTAGGC CCGGGACCAAG CGGTCTCGGT
 4141 CGTGAGGGT CCGTGTATT TTTTCCAGGA CGTGGTAAAG GTGACTCTGG ATGTTCAGAT
 4201 ACATGGGCAT AAGCCCGTCT CTGGGGTGGG GGTAGCACCA CTGAGAGCT TCATGCTGCG
 4261 GGGTGGTGTG GTAGATGATC CAGTCGTAGC AGGAGCGCTG GGCCTGGTGC CTAAAAATGT
 4321 CTTTCAGTAG CAAGCTGATT GCCAGGGCA GGCCCTTGGT GTAAAGTGT ACAAAGCGGT
 4381 TAAGCTGGGA TGGGTGCATA CGTGGGGATA TGAGATGCAT CTTGGACTGT ATTTTTAGGT
 4441 TGGCTATGTT CCCAGCCATA TCCCTCCGGG GATTCTATGTT GTGCAGAAC ACCAGCACAG
 4501 TGTATCCGGT GCACCTGGGA AATTGTCAT GTAGCTTAGA AGGAAATGCG TGGAAGAACT
 4561 TGGAGACGCC CTTGTGACCT CCAAGATTT CCATGCATTC GTCCATAATG ATGGCAATGG
 4621 GCCCACGGGC GGCGGCCTGG CGAAGATAT TTCTGGGATC ACTAACGTCA TAGTTGTGTT
 4681 CCAGGATGAG ATCGTCATAG GCCATTTTA CAAAGCGCGG CGGGAGGGTG CCAGACTGCG
 4741 GTATAATGGT TCCATCCGGC CCAGGGCGT AGTTACCCCTC ACAGATTGCG ATTTCACCG
 4801 CTTTGAGTTC AGATGGGGGG ATCATGTCAT CCTGCGGGGC GATGAAGAAA ACGGTTTCCG
 4861 GGGTAGGGGA GATCAGCTGG GAAGAAAGCA GGTTCCTGAG CAGCTGCAGC TTACCGCAGC
 4921 CGGTGGGCC CTAATCACA CCTATTACCG GGTGCAACTG GTAGTTAAGA GAGCTGCAGC
 4981 TGCCGTATC CCTGAGCAGG GGGGCCACTT CGTAAAGCAT GTCCCTGACT CGCATGTTT
 5041 CCCTGACCAA ATCCGCCAGA AGGCCTCGC CGCCCAAGCGA TAGCTGCAGC TTACCGCAGC
 5101 CAAAGTTTTT CAACGGTTTG AGACCGTCCG CGTAGGCAT GCTTTTGAGC GTTTGACCAA
 5161 GCAGTCCAG GCGGTCCCAC AGCTCGGTCA CCTGCTCTAC GGCACTCGA TCCAGCATAT
 5221 CTCCTCGTT CGCGGGTTGG GGGCGCTTC GCTGTAACGGC AGTAGTCGGT GCTCGTCCAG
 5281 ACGGGCCAGG GTCATGTCCT TCCACGGCG CAGGGCTCTC GTCACTGCAG TCTGGGTAC
 5341 GGTGAAGGGG TCGCGCTCCGG GCTGCGCGT GGCCAGGGTG CGCTTGAGGC TGGTCTGCT
 5401 GGTCGCTGAAG CGCTGCCGGT CTTCGCCCCG CGCGCTCGGCC AGGTAGCATT TGACCATGGT
 5461 GTCATAGTCC AGGCCCTCCG CGGGCTGGC CTTGGCGCGC AGCTTGCCT TGGAGGAGGC
 5521 GCGCACGAG GGGCAGTGCA GACTTTTGAG GGCCTAGAGC TTGGGCGCGA GAAATACCGA
 5581 TTCCGGGGAG TAGGCATCCG CGCCGCAAGC CCCCAGACG GTCTCGCATT CCACGAGCCA
 5641 GGTGAGGCTCT GGGCGTTCCG GGTCAAAAC CAGGTTTCCC CCATGCTTT TGATGCGTTT
 5701 CTTACCTCTG GTTCCATGA GCGGTGTCC ACGCTCGGTG ACAGAAAAGGC TGTCCGTGTC
 5761 CCCGTATACA GACTTGAGAG GCCTGTCCTC GAGCGGTGTT CGCGGGCTCT CTCGTATAG
 5821 AAACTCGGAC CACTCTGAGA CAAAGGCTCG CGTCCAGGCC AGCACGAAGG AGGCTAAGTG
 5881 GGAGGGGTAG CGGTGCGTTGT CCACTAGGGG GTCCACTCGC TCCAGGGTGT GAAGACACAT
 5941 GTGCCCTCT TCAGCATCAA GGAAGGTGAT TGGTTGAG GTGTAGGCCA CGTGACCGGG
 6001 TGTTCTGAA GGGGGGCTAT AAAAGGGGT GGGGGCGCGT TCGTCTCAG TCTCTCCGC
 6061 ATCGCTGTCT GCGAGGGCCA GCTGTTGGGG TGAGTACTCC CTCTGAAAAG CGGGCATGAC
 6121 TTCTGCGCTA AGATTGTCAG TTTCCAAAAA CGAGGAGGAT TTGATATTCA CCTGGCCCGC
 6181 GGTGATGCGCT TTGAGGGTGG CGGCATCCAT CTGGTCAGAA AAGACAATCT TTTTGTGTC
 6241 AAGCTGGTG GCAAACGACC CGTAGAGGGC GTTGGACAGC AACCTGGCGA TGGAGCGCAG
 6301 GGTTGGTTT TTGTCGCGAT CGGCGCCCTC CTGGCCCGCG ATGTTTAGCT GCACGTATT
 6361 GCGCGCAACG CACCGCCATT CGGGAAAGAC GGTGGTGCAGC TCGTCGGGCA CCAGGTGCAC
 6421 GCGCCAACCG CGGTGCGCA CGGTGACAAG GTCAACGCTG GTGGCTACCT CTCCCGTAG
 6481 GCGCTCGTTG GTCCAGCAGA GGCGGCCGCC CTTGCGCGAG CAGAATGGCG GTAGGGGGTC
 6541 TAGCTGCGTC TCGTCCGGGG GGTCTGCGTC CACGGTAAAG ACCCCGGGCA CGAGGCGCGC

59/92

6601 GTCGAAGTAG TCTATCTTGC ATCCTTCAA GTCTAGCGCC TGCTGCCATG CGCGGGCGGC
 6661 AAGCGCGCGC TCGTATGGGT TGAGTGGGG ACCCCATGGC ATGGGGTGGG TGAGCGCGGA
 6721 GCGTACATG CCGCAAATGT CGTAAACGTA GAGGGGCTCT CTGAGTATTG CAAGATATGT
 6781 AGGGTAGCAT CTTCACCAGC GGATGCTGGC GCGCACGTA TCGTATAAGTT CGTGCAGGG
 6841 AGCGAGGAGG TCGGGACCGA GGTTGCTACG GGCGGGCTGC TCTGCTCGGA AGACTATCTG
 6901 CCTGAAGATG GCATGTGAGT TGGATGATAT GGTTGGACGC TCCAAGACGT TGAAGCTGGC
 6961 GTCTGTGAGA CCTACCGCGT CACGCACGAA GGAGGCGTAG GAGTCGCGCA GCTTGTGAC
 7021 CAGCTCGCGC GTGACCTGCA CGTCTAGGGC GCAGTAGTCC AGGGTTTCCCT TGATGATGTC
 7081 ATACTTATCC TGTCCTTTT TTTTCCACAG CTCGCCTTG AGGACAAACT CTTCGCGGTC
 7141 TTTCCAGTAC TCTTGGATCG GAAACCGTC GGCTCCGAA CGGTAAGAGC CTAGCATGTA
 7201 GAACTGGTTG ACGGCCTGGT AGGCGCAGCA TCCCCTTCT ACGGGTAGCG CGTATGCCTG
 7261 CGCGCCCTTC CGGAGCGAGG TGTGGGTGAG CGCAAAGGTG TCCCTGACCA TGACTTTGAG
 7321 GTACTGGTAT TTGAAGTCAG TGTCGTCGA TCCGCCCTGC TCCCAGAGCA AAAAGTCCGT
 7381 GCGCTTTTG GAACGCGGGAT TTGGCAGGGC GAAGGTGACA TCGTTGAAGA GTATCTTCC
 7441 CGCGCGAGGC ATAAGGTTGC GTGTGATGCG GAAGGGTCCC GGCACCTCGG AACGGTTGTT
 7501 AATTACCTGG GCGGCGAGCA CGATCTCGTC AAAGCCGTTG ATGTTGTGCG CCACAATGTA
 7561 AAGTTCCAAG AAGCGCGGGA TGCCCTTGAT GGAAGGCAAT TTTTTAAGTT CCTCGTAGGT
 7621 GAGCTCTTC GGGGAGCTGA GCCCCGTGCT TGAAAGGGCC CAGTCTGCAA CATGAGGGTT
 7681 GGAAGCGACG AATGAGCTCC ACAGGTACG GGCACATTAGC ATTTGCAGGT GGTCCGAAA
 7741 GGTCTAACAC TGGCGACCTA TGGCCATTTC TTCTGGGTG ATGCAGTAGA AGGTAAGCGG
 7801 GTCTGTTC CAGCGGTCCC ATCCAAGGTT CGCCGCTAGG TCTCGCGCGG CAGTCACTAG
 7861 AGGCTCATCT CCGCGGAAC TCATGACCAG CATGAAGGGC ACGAGCTGCT TCCCAAAGGC
 7921 CCCCATCCAA GTATAGGTCT CTACATCGTA GGTGACAAAG AGACGCTCGG TGCAGGGATG
 7981 CGAGCCGATC GGGAGAAACT GGATCTCCG CCACCAATTG GAGGAGTGGC TATTGATGTC
 8041 GTGAAAGTAG AAGCTCTGC GACGGGCGA ACACTCGTC TGGCTTTGT AAAAACGTGC
 8101 GCAGTACTGG CAGCGGTGCA CGGGCTGTAC ATCCTGCAAG AGGTTGACCT GACGACCGCG
 8161 CACAAGGAAG CAGAGTGGGA ATTTGAGCCC CTCCGCTGGC GGGTTGGCT GGTGGCTTTC
 8221 TACTTCGGCT GCTTGTCTT GACCGTCTGG CTGCTCGAGG GGAGTTACGG TGGATCGGAC
 8281 CACCACGCCG CGCGAGGCCA AAGTCAGAT GTCCGCGGGC GCGGGTCCGA GCTTGATGAC
 8341 AACATCGCGC AGATGGGAGC TGTCCATGGT CTGGAGCTCC CGCGCGTCA GGTCAAGGGG
 8401 GAGCTCTGC AGGTTTACCT CGCATAGACG GGTCAAGGGCG CGGGCTAGAT CCAGGTGATA
 8461 CCTAATTTCG AGGGGCTGGT TGGTGGGGC GTCCGATGGCT TGCAAGAGGC CGCATCCCCG
 8521 CGGCGCGACT ACGGTACCGC CGGGGGGGC GTGGGCCCCG GGGGTGTCTT TGGATGATGC
 8581 ATCTAAAAGC GGTGACCGGG CGAGCCCCC GGAGGTAGGG GGGGCTCCGG ACCCGCCGGG
 8641 AGAGGGGCA GGGGCACGTC GGCACCGCGC GCGGGCAGGA GCTGGTGCTG CGCGCGTAGG
 8701 TTGCTGGCGA ACGCGACGAC GCGGCGTTG ATCTCCTGAA TCTGGCGCCT CTGCGTGAAG
 8761 ACGACGGCC CGGTGAGCTT GAGCCTGAA GAGAGTTCGA CAGAATCAAT TTGGTGTGCG
 8821 TTGACGGCGG CCTGGCGAA AATCTCTGC ACGTCTCTG AGTTGTCTG ATAGGCGATC
 8881 TCGGCCATGA ACTGCTCGAT CTCTTCTCTC TGGAGATCTC CGCGTCCGGC TCGCTCCACG
 8941 GTGGCGCGA GGTGTTGGA AATGCGGGCC ATGAGCTGCG AGAAGGCCTT GAGGCCCTCCC
 9001 TCGTTCCAGA CGCGCGCTGA GACCACGCC CCTTCGGCAT CGCGGGCGCG CATGACCAAC
 9061 TCGCGAGAT TGAGCTCAC GTGCCGGCG AAGACGGCGT AGTTTCGAG GCGCTGAAAG
 9121 AGGTAGTTGA GGGTGGTGGC GGTGTGTTCT GCCACGAAGA AGTACATAAC CCAGCGTCGC
 9181 AACGTGGATT CGTTGATATC CCCCAGGCC TCAAGGCGCT CCATGGCCCTC GTAGAAAGTCC
 9241 ACGGCGAAGT TGAAGAAACTG GGAGTTGCGC GCCGACACGG TTAACTCCCTC CTCCAGAAGA
 9301 CGGATGAGCT CGCGACAGT GTCGCGCAC TCGCGCTCAA AGGCTACAGG GGCCTCTTCT
 9361 TCTTCTCAA TCTCTCTTC CATAAGGGCC TCCCCCTCTT CTTCTCTGG CGCGGGTGGG
 9421 GGAGGGGGCA CACGGCGCGC ACGACGGCGC ACCGGGAGGC GGTCGACAAA GCGCTCGATC
 9481 ATCTCCCCGC CGCGACGGCG CATGGTCTCG GTGACGGCGC GGCGCTCTC CGGGGGCGC
 9541 AGTTGGAAGA CGCCGCGCGT CATGTCCCGG TTATGGGTG GCGGGGGGCT GCCATGCGGC
 9601 AGGGATACGG CGCTAACGAT GCATCTCAAC AATTGTTGTG TAGGTACTCC GCGCGCGAGG
 9661 GACCTGAGCG AGTCCGCACTC GACCGGATCG GAAAACCTCT CGAGAAAGGC GTCTAACAG
 9721 TCACAGTCGC AAGGTAGGCT GAGCACCGTG GCGGGCGGCA GCGGGCGGGC GTCGGGGTTG
 9781 TTTCTGGCGG AGGTGCTGCT GATGATGTAA TTAAAGTAGG CGGTCTTGAG ACGGCGGATG
 9841 GTCGACAGAA GCACCATGTC CTTGGTCCG GCCTGCTGAA TGCGCAGGGC GTCGGCCATG

FIG. 8C

60/92

9901 CCCCAGGCTT CGTTTGACA TCGGCGCAGG TCTTTGTAGT AGTCTTGCAT GAGCCTTCT
 9961 ACCGGCACTT CTTCTCTCC TTCCCTTGT CCTGCATCTC TTGCATCTAT CGCTGCGGCG
 10021 GCGGCGGAGT TTGGCCGTAG GTGGCGCCCT CTTCTCCCA TGCGTGTGAC CCCGAAGCCC
 10081 CTCATCGGCT GAAGCAGGGC TAGGTCGGC ACAACGCGCT CGGCTAATAT GGCGCTGCTGC
 10141 ACCTGCGTGA GGGTAGACTG GAAGTCATCC ATGTCACAA AGCGGTGGTA TGCGCCCGTG
 10201 TTGATGGTGT AAGTGCAGTT GGCCATAACG GACCAGTTAA CGGTCTGGTG ACCCGGCTGC
 10261 GAGAGCTCGG TGTACCTGAG AC CGGAGTAA GCCCTCGAGT CAAATACGTA GTCGTTGCAA
 10321 GTCCCGACCA GGTACTGGTA TCCCCACAAA AAGTGCAGGC GCGGCTGGCG GTAGAGGGGC
 10381 CAGCGTAGGG TGGCGGGGC TCCGGGGCG AGATCTCCA ACATAAGGCC ATGATATCCG
 10441 TAGATGTACC TGGACATCCA GGTGATGCCG GCGGCGGTGG TGGAGGCCG CGGAAAGTCG
 10501 CGGACGCGGT TCCAGATGTT GCGCAGCCGC AAAAAGTGCT CCATGGTCGG GACGCTCTGG
 10561 CGCGTCAGGC CGCGCGCAATC GTTGCACGCTC TAGACCGTGC AAAAGGAGAG CCTGTAAGCG
 10621 GGCACCTTC CGTGGCTCTGG TGGATAAATT CGCAAGGTA TCATGGCGGA CGACGGGGT
 10681 TCGAGCCCCG TATCCGGCCG TCCGGCGTGA TCCATGCGGT TACCGCCCGC GTGTCGAACC
 10741 CAGGTGTGCG AC GTCAGACA AC GGGGGAGT GCTCCTTTG GCTTCCTTCC AGGCGCGGCG
 10801 GCTGCTGCGC TAGCTTTTTT GGCCACTGGC CGCGCGCAGC GTAAGCGGTT AGGCTGGAAA
 10861 GCGAAAGCAT TAAGTGGCTC GCTCCCTGTA GCCGGAGGGT TATTTCCAA GGGTTGAGTC
 10921 GCGGGACCCC CGGTTCGAGT CTCGGACCGG CGGACTGCG GCGAACGGGG GTTTGCCCTCC
 10981 CCGTCATGCA AGACCCCGCT TGCAAATCC TCCGAAACAA GGGACGAGCC CCTTTTTTGC
 11041 TTTTCCCAAGA TGATCCGGT GCTGCGGAG ATGCGCCCCC CTCCCTCAGCA CGGGCAAGAG
 11101 CAAGAGCAGC GGCAGACATG CAGGGCACCC TCCCCTCCTC CTACCGCGTC AGGAGGGCG
 11161 ACATCCCGGG TTGACGCGGC AGCAGATGGT GATTACGAAC CCCC CGCGGCG CGGGGGCCCG
 11221 CACTACCTGG ACTTGGAGGA GGGCGAGGGC CTGGCGCGGC TAGGAGCGCC CTCTCCTGAG
 11281 CGGTACCCAA GGGTGCAGCT GAAGCGTGAT ACGCGTGAGG CGTACGTGCC CGGGCAGAAC
 11341 CTGTTTCGCG ACCCGAGGG AGAGGAGGCC GAGGAGATGC GGGATCGAAA GTTCCACGCA
 11401 GGGCGCGAGC TGGCGCATGG CCTGAATCGC GAGCGGTG TGCGCGAGGA GGACTTTGAG
 11461 CCCGACGCGC GAACCGGGAT TAGTCCCGCG CGCGCACACG TGGCGGCCCG CGACCTGGTA
 11521 ACCGCATACG AGCAGACGGT GAACCAGGAG ATTAACTTC AAAAAAGCTT TAACAACCAC
 11581 GTCGTGTACGC TTGTGGCGCG CGAGGAGGTG GCTATAGGAC TGATGCATCT GTGGGACTTT
 11641 GTAAGCGCGC TGAGCAAAA CCCAAATAGC AAGCCGCTCA TGGCGCAGCT GTTCCTTATA
 11701 GTGCAGCACCA GCAGGGACAA CGAGGCATTC AGGGATGCGC TGCTAAACAT AGTAGAGGCC
 11761 GAGGGCCCGT GGCTGCTCGA TTGATAAAC ATCCTGCAGA GCATAGTGGT GCAGGAGCGC
 11821 AGCTTGAGCC TGCTGACAA GGTGGCCGCC ATCAACTATT CCATGCTTAG CCTGGCAAG
 11881 TTTTACGCC GCAAGATATA CCATACCCCT TACGTTCCA TAGACAAGGA GTAAAGATC
 11941 GAGGGGTTCT ACATGCGCAT GGC GCTGAAG GTGCTTACCT TGAGCGACCA CCTGGCGTT
 12001 TATCGCAACG AGCGCATCCA CAAGGCGGTG AGCGTGAGCC GGC GCGCGCA GCTCAGCGAC
 12061 CGCGAGCTGA TGCACAGCCT GCAAAGGCC CTGGCTGGCA CGGGCAGCGG CGATAGAGAG
 12121 GCCGAGTCCT ACTTTGACGC GGGCGCTGAC CTGGCTGGG CCCCAGGCC AC GCGCCCTG
 12181 GAGGCAGCTG GGGCGGGACC TGGGCTGGCG GTGGCACCCCG CGCGCGCTGG CAACGTCGGC
 12241 GGC GTGGAGG AATATGACGA GGACGATGAG TACCGAGCCAG AGGACGGCGA GTACTAAGCG
 12301 GTGATGTTTC TGATCAGATG ATGCAAGACG CAACGGACCC GGCGGTGGCG GCGGCGCTGC
 12361 AGAGCCAGCC GTCCGGCCTT AACTCCACGG ACGACTGGCG CCAGGTCATG GACCGCATCA
 12421 TGTCGCTGAC TGC GCGCAAT CCTGACGCGT TCCGGCAGCA GCGCAGGCC AACCGGCTCT
 12481 CCGCAATTCT GGAAGCGGTG GTCCCGGGCG GCGCAAACCC CACGCACGAG AAGGTGCTGG
 12541 CGATCGTAA CGCGCTGGCC GAAAACAGGG CCATCCGGCC CGACGAGGCC GGCCTGGTCT
 12601 ACGACCGCCT GCTTCAGCGC GTGGCTCGTT ACAACAGCGG CAACGTGCAG ACCAACCTGG
 12661 ACCGGCTGGT GGGGGATGTG CGCGAGGCCG TGGCGCAGCG TGAGCGCGC CAGCAGCAGG
 12721 GCAACCTGGG CTCCATGGTT GCACTAAACG CCTTCCTGAG TACACAGCCC GCCAACGTGC
 12781 CGCGGGACCA GGAGGACTAC ACCAACTTT TGAGCGCACT GC GGCTAATG GTGACTGAGA
 12841 CACCGCAAAG TGAGGTGTAC CAGTCTGGC CAGACTATTT TTTCCAGACC AGTAGACAAG
 12901 GCCTGCGACAC CGTAAACCTG AGCCAGGCTT TCAAAACATT GCAGGGCTG TGGGGGTGCG
 12961 GGGCTCCAC AGGCGACCGC GCGACCGTGT CTAGCTTGCT GACGCCAAC TCGCGCCTGT
 13021 TGCTGCTGCT AATAGCGCCC TTCACGGACA GTGGCAGCGT GTCCCGGGAC ACATAACCTAG
 13081 GTCACTTGCT GACACTGTAC CGCGAGGCCA TAGGTCAGGC GCATGTGGAC GAGCATACTT
 13141 TCCAGGAGAT TACAAGTGTCA AGCCCGCGC TGGGGCAGGA GGACACGGC AGCCTGGAGG

FIG. 8D

13201 CAACCCCTAAA CTACCTGCTG ACCAACCGC GGCAGAAAGAT CCCCTCGTTG CACAGTTAA
 13261 ACAGCGAGGA GGAGCGCATT TTGCGCTACG TGCAGCAGAG CGTGAGCCTT AACCTGATGC
 13321 GCGACGGGGT AACGCCAGC GTGGCGCTGG ACATGACCAGC GCGCAACATG GAACCGGGCA
 13381 TGTATGCCCTC AAACCGGCGC TTTATCAACC GCCTAATGGA CTACTTGCT CGCGCGGCCG
 13441 CCGTGAACCC CGAGTATTTG ACCAATGCCA TCTTGAAACC GCACTGGCTA CCGCCCCCTG
 13501 GTTTCCTACAC CGGGGGATTC GAGGTGCCCG AGGTAACGA TGGATTCTC TGGGACGACA
 13561 TAGACGACAG CGTGTGTTTC CCGCAACCAGC AGACCCCTGCT AGAGTTGCAA CAGCGCGAGC
 13621 AGGCAGAGGC GGCCTGCGCA AAGGAAGCT TCCGCAGGCC AAGCAGCTTG TCCGATCTAG
 13681 GCGCTGCGGC CCCCGGGTCA GATGCTAGTA GCCCATTTCC AAGCTTGATA GGGTCTCTTA
 13741 CCAGCACTCG CACCACCCGC CCGCGCTGC TGCGCAGGAGA GGAGTACCTA AACAACTCGC
 13801 TGCTGCAGCC GCAGCGCGAA AAAAACCTGC CTCCGGCATT TCCCAACAAAC GGGATAGAGA
 13861 GCCTAGTGGA CAAGATGAGT AGATGGAAGA CGTACCGCAGA GGAGCACAGG GACGTGCCAG
 13921 GCCCCGGCCC GCCCACCCGT CGTCAAAGGC ACGACCGTCA GCGGGGCTCG GTGTTGGAGG
 13981 ACGATGACTC GGCAGACGAC AGCAGCGTCC TGGATTGGG AGGGAGTGGC AACCGTTTG
 14041 CGCACCTTCG CCCCAGGCTG GGGAGAATGT TTTAAAAAAA AAAAACGATG ATGCAAAATA
 14101 AAAAACCTCAC CAAGGCCATG GCACCGAGCG TTGGTTTCT TGTATTCCCC TTAGTATGCG
 14161 GCGCGGGCG ATGTATGAGG AAGGTCCTCC TCCCTCTAC GAGAGTGTG TGACGCCGGC
 14221 GCCAGTGGCG GCGGGCGCTGG GTTCTCCCTT CGATGCTCCC CTGGACCCGC CGTTTGTGCC
 14281 TCCCGGGTAC CTGCGGCCCTA CGGGGGGGAG AAACAGCATT CGTTACTCTG AGTTGGCAC
 14341 CCTATTCGAC ACCACCCGTG TGTACCTGGT GGACAACAAG TCAACGGATG TGGCATCCCT
 14401 GAACTACCAAG AACGACCCACA GCAACTTCT GACACCGGT ATTCAAAACA ATGACTACAG
 14461 CCCGGGGAG GCAAGCACAC AGACCATCAA TCTGACGAC CGGTCGACT GGGCGCGA
 14521 CCTGAAAACC ATCCTGCTA CCAACATGCC AAATGTGAAC GAGTTCATGT TTACCAATAA
 14581 GTTTAAGGCG CGGGTGATGG TGTCGCGCTT GCCTACTAAG GACAATCAGG TGGAGCTGAA
 14641 ATACGACTGG GTGGAGTTCA CGCTGCCGA GGGCAACTAC TCCGAGACCA TGACCATAGA
 14701 CCTTATGAAC AACCGATCG TGGAGCACTA CTTGAAAGTG GGCAGACAGA ACGGGGTTCT
 14761 GGAAAGCGAC ATCAGGGTAA AGTTTGACAC CGCAACTTC AGACTGGGGT TTGACCCCGT
 14821 CACTGGTCTT GTCATGCCCTG GGGTATATAC AAACGAAGCC TTCCATCCAG ACATCATT
 14881 GCTGCCAGGA TGCGGGGTGG ACTTCACCCA CAGCCGCCCT AGCAACTTGT TGGGCATCCG
 14941 CAAGCGGCAA CCTTCCAGG AGGGCTTTAG GATCACCTAC GATGATCTGG AGGGTGGTAA
 15001 CATTCCCGCA CTGTTGGATG TGGACGCCCTA CCAGGGGAGC TTGAAAGATG ACACCGAAC
 15061 GGGCGGGGGT GGCAGAGGC GCAGCAACAG CAGTGGCAGC GGCGCGGAAG AGAACTCCAA
 15121 CGCGCAGCC GCGCAATGC AGCCGGTGGA GGACATGAAC GATCATGCCA TTCCGCCGA
 15181 CACCTTGCC ACACGGGCTG AGGAGAAGCG CGCTGAGGCC GAAGCAGCGG CCGAAGCTGC
 15241 CGCCCCCGCT GCGCAACCCCG AGGTCGAGAA GCCTCAGAAG AAACCGGTGA TCAAACCCCT
 15301 GACAGAGGAC AGCAAGAACAC GCACTTACAA CCTAATAAGC AATGACAGCA CCTTCACCC
 15361 GTACCGCAGC TGTTACCTTG CATACAACTA CGGGCACCC CAGACCGGAA TCCGCTCATG
 15421 GACCCCTGCTT TGCACTCCCTG ACGTAACCTG CGGCTCGGAG CAGGTCTACT GGTCTTGCC
 15481 AGACATGATG CAAGACCCCG TGACCTTCCG CTCCACGCC CAGATCAGCA ACTTCCGGT
 15541 GGTGGCGCC GAGCTGTTGC CGTGCACTC CAAGAGCTTC TACAACGACC AGGCCGTCTA
 15601 CTCCCAACTC ATCCGCCAGT TTACCTCT GACCCACGT TTCATCGCT TTCCCGAGAA
 15661 CCAGATTTG GCGCGCCCGC CAGCCCCAC CATACCAAC GTCAGTAAAA ACGTCCCTGC
 15721 TCTCACAGAT CACGGGACGC TACCGCTGCG CAACAGCATT GGAGGAGTCC AGCGAGTGAC
 15781 CATTACTGAC GCCAGACGCC GCACCTGCCCT CTACGTTAC AAGGCCCTGG GCATAGTCTC
 15841 GCCGCCGTC CTATCGAGCC GCACCTTTG AGCAAGCATG TCCATCCTTA TATGCCAG
 15901 CAATAACACA GGCTGGGGCC TGCGCTTCCC AAGCAAGATG TTGGCGGGG CCAAGAAGCG
 15961 CTCCGACCAA CACCCAGTGC GCGTGCAGCG GCACTACCAGC GCGCCCTGGG GCGCCACAA
 16021 ACGCGGCCGC ACTGGGCGCA CCACCGTCGA TGACGCCATC GACCGGGTGG TGGAGGAGGC
 16081 GCGCAACTAC ACGCCCACGC CGCCACCAAGT GTCCACAGTG GACCGGGCCA TTCAAGACCGT
 16141 GGTGGCGGA GCGCGCCGCT ATGCTAAAT GAAGAGACGG CGGAGGCGCG TAGCACGTCG
 16201 CCACCCCGC CGACCCGGCA CTGCCGCCA ACAGCGCGCG GCGCCCTGC TTAACCGCGC
 16261 ACGTCCGACC GGGCGACGGG CGGCCATGCG GGCGCTCGA AGGCTGGCC CGGGTATTGT
 16321 CACTGTGCCCTC CCCAGGTCCA GGCAGCAGC GGCGCCGCA GCAGCCGCGG CCATTAGTGC
 16381 TATGACTCAG GTGCGCAGGG GCAACGTGTA TTGGGTGCGC GACTCGGTTA GCGGCCCTGCG
 16441 CGTCCCCGTG CGCACCCGCC CCCCCGCGAA CTAGATTGCA AGAAAAAAACT ACTTAGACTC

62/92

16501 GTACTGTTGT ATGTATCCAG CGGCAGCGGC GCGCAACGAA GCTATGTCCA AGCGCAAAT
 16561 CAAAGAAGAG ATGCTCCAGG TCATCGGCC GGAGATCTAT GGCCCCCGA AGAAGGAAGA
 16621 GCAGGATTAC AAGCCCCGAA AGCTAAAGCG GGTCAAAAAG AAAAGAAAG ATGATGATGA
 16681 TGAACGTGAC GACGGAGGTGG AACTGCTGCA CGCTACCGCG CCCAGGCGAC GGGTACAGTG
 16741 GAAAGGTCGA CGCGTAAAAC GTGTTTGC ACCCGGCACC ACCGTAGTCT TTACGCCGG
 16801 TGAGCGCTCC ACCCGCACCT ACAAGCGCTG GTATGATGAG GTGTACGGCG ACGAGGACCT
 16861 GCTTGAGCAG GCCAACGAGC GCCTCGGGGA GTTTGCCTAC GGAAAGCGGC ATAAGGACAT
 16921 GCTGGCGTTG CCGCTGGACG AGGGCAACCC AACACCTAGC CTAAAGCCCC TAACACTGCA
 16981 GCAGGGTCTG CCCCGCGTTG CACCGTCCGA AGAAAAGCGC GGCCTAAAGC GCGAGTCTGG
 17041 TGACTTGGCA CCCACCGTGC AGCTGATGGT ACCCAAGCGC CAGCGACTGG AAGATGTCTT
 17101 GGAAAAAATG ACCGTGGAAC CTGGGCTGGA GCCCCGAGGTC CGCGTGCAGC CAATCAAGCA
 17161 GGTGGCGCCG GGACTGGGCG TGCAAGACCGT GGACGTTCAAG ATACCCACTA CCAGTAGCAC
 17221 CAGTATTGCC ACCGCCACAG AGGGCATGGA GACACAAACG TCCCCGGTTG CCTCAGCGGT
 17281 GGCAGATGCC GCGGTGCAGG CGGTCGCTGC GGCGCGTCC AAGACCTCTA CGGAGGTGCA
 17341 AACGGACCCG TGGATGTTTC GCGTTTCAGC CCCCCCGCGC CGCGCGGGT CGAGGAAGTA
 17401 CGGCGCCGCC AGCGCGCTAC TGCCCCATA TGCCCTACAT CCTTCCATTG CGCCTACCCC
 17461 CGGCTATCGT GGCTACACCT ACCGCCCGAG AAGACGAGCA ACTACCCGAC GCGAACAC
 17521 CACTGGAACC CGCCGCCGCC GTCGCCGTGC CCAGCCCGTG CTGGCCCCGA TTTCCGTGCG
 17581 CAGGGTGGCT CGCGAAGGAG GCAGGACCCCT GGTGCTGCCA ACAGCGCGCT ACCACCCAG
 17641 CATCGTTAA AAGCCGGTCT TTGTGGTTCT TGCAAGATATG GCCCTCACCT GCGCCTCCG
 17701 TTTCCCGGTG CCGGGATTCC GAGGAAGAAT GCACCGTAGG AGGGCCATGG CGGGCCACGG
 17761 CCTGACGGGC GGCATGCGTC GTGCGCACCA CCGCGGGCGG CGCGCGTCCG ACCGTCGCAT
 17821 GCGCGCGGT ATCCTGCCCG TCCTTATTCC ACTGATCGCC GCGGCATGG GCGCCGTGCC
 17881 CGGAATTGCA TCCGTGGCCT TGCAAGCGA GAGACACTGA TAAAAAAACAA GTTGCATGTG
 17941 GAAAAATCAA AATAAAAAGT CTGGACTCTC ACGCTCGCTT GGTCTGTAA CTATTTGTA
 18001 GAATGGAAGA CATCAACTTT GCGTCTCTGG CCCCCGCGACA CGGCTCGCC CGGTTCATGG
 18061 GAAACTGGCA AGATATCGGC ACCAGCAATA TGAGCGGTGG CGCCTTCAGC TGGGGCTCGC
 18121 TGTGGAGCGG CATTAAAAAT TTCGGTTCCA CCGTTAAGAA CTATGGCAGC AAGGCTGGAA
 18181 ACAGCAGCAC AGGCCAGATG CTGAGGGATA AGTTGAAAGA GAAAAATTTC CAACAAAAGG
 18241 TGGTAGATGG CCTGGCTCTGG GGCATTAGCG GGGTGGTGGG CCTGGCCAAC CAGGCAGTGC
 18301 AAAATAAGAT TAACAGTAAG CTTGATCCCC CGCCCTCCCGT AGAGGAGCCT CCACCGGCCG
 18361 TGGAGACAGT GTCTCCAGAG GGGCGTGGCG AAAAGCGTCC GCGCCCCGAC AGGGAGGAAA
 18421 CTCTGGTGCAC GCAAATAGAC GAGCCTCCCT CGTACCGAGGA GGCACTAAAG CAAGGCTGC
 18481 CCACCAACCG TCCCATCGCG CCCATGGCTA CGGAGGTGCT GGGCCAGCAC ACACCCGTAA
 18541 CGCTGGACCT GCTCCCCCCC GCGCACACCC AGCAGAAACC TGTGCTGCCA GGCCCCGACCG
 18601 CGGTTGTTGTG AACCCTGCTCT AGCCGCGCTG CCCTCGCGCG CGCGGCCAGC GGTCCCGCAT
 18661 CGTTGGGGCC CGTAGCCAGT GGCAACTGGC AAAGCACACT GAACAGCATC GTGGGTCTGG
 18721 GGGTGAATC CCTGAAGCGC CGACGATGCT TCTGAATAGC TAACGTGTCG TATGTGTGTC
 18781 ATGTATGCGT CCATGTCGCC GCCAGAGGAG CTGCTGAGCC GCCGCGCGCC CGCTTCCAA
 18841 GATGGCTACC CCTTCGATGA TGCCCGAGTG GTCTTACATG CACATCTCGG GCCAGGACGC
 18901 CTCGGAGTAC CTGAGCCCCG GGCTGGTGCA GTTGTCCCAGC GCCACCGAGA CGTACTTCAG
 18961 CCTGAATAAC AAGTTTAGAA ACCCCACGGT GGCCCTACG CACGACGTGA CCACAGACCG
 19021 GTCCCAGCGT TTGACGCTGC GGTTCATCCC TGTGGACCGT GAGGATACTG CGTACTCGTA
 19081 CAAGGGCGGG TTACCCCTAG CTGTGGGTGA TAACCGTGTG CTGGACATGG CTTCCACGTA
 19141 CTTTGACATC CGCGCGTGC TGGACAGGGG CCCTACTTTT AAGCCCTACT CTGGCACTGC
 19201 CTACAACGCC CTGGCTCCCA AGGGTCCCCC AAATCCTGC GAATGGGATG AAGCTGCTAC
 19261 TGCTCTTGAAT AAAACCTAG AAGAAGAGGA CGATGACAAC GAAGACGAAG TAGACGAGCA
 19321 AGCTGAGCAG CAAAAAAACTC ACGTATTGAG GCAGGCGCCT TATTCTGGTA TAAATATTAC
 19381 AAAGGAGGGT ATTCAAATAG GTGTCGAAGG TCAAAACACCT AAATATGCCG ATAAAACATT
 19441 TCAACCTGAA CCTCAAATAG GAGAATCTCA GTGGTACGAA ACTGAAATTA ATCATGCGAC
 19501 TGGGAGAGTC CTAAAAAGA CTACCCCAAT GAAACCATGT TACGGTTCAT ATGCAAACCC
 19561 CACAAATGAA AATGGGAGGGC AAGGCATTCT TGTAAGCAA CAAATGGAA AGCTAGAAAAG
 19621 TCAAGTGGAA ATGCAATTTC TCTCAACTAC TGAGGCGACC GCAGGCAATG GTGATAACTT
 19681 GACTCCTAAA GTGGTATTGT ACAGTGAAGA TGTAGATATA GAAACCCAG ACACTCATAT
 19741 TTCTTACATG CCCACTATTA AGGAAGGTAA CTCACGAGAA CTAATGGGCC AACAAATCTAT

FIG. 8F

19801 GCCCAACAGG CCTAATTACA TTGCTTTAG GGACAATTT ATTGGTCTAA TGTATTACAA
 19861 CAGCACGGGT AATATGGGT TTCTGGGGG CCAAGCATCG CAGTTGAATG CTGTTGTAGA
 19921 TTTGCAAGAC AGAAAACACAG AGCTTTCATA CCAGCTTTG CTTGATTCCA TTGGTGTAGA
 19981 ACCCAGGTAC TTTTCTATGT GGAATCAGGC TGTTGACAGC TATGATCCAG ATGTTAGAAT
 20041 TATTGAAAAT CATGGAACTG AAGATGAACT TCCAAATTAC TGCTTCCAC TGGGAGGTGT
 20101 GATTAATACA GAGACTCTTA CCAAGGTAAA ACCTAAAACA GGTCAGGAAA ATGGATGGGA
 20161 AAAAGATGCT ACAGAATTTC CAGATAAAAA TGAAATAAGA GTTGGAAATA ATTTTGCCAT
 20221 CGAAATCAAT CTAATGCCA ACCTGTGGAG AAATTCCTG TACTCCAACA TAGCGCTGTA
 20281 TTGCCCCGAC AAGCTAAAGT ACAGTCCTTC CAACGTAAAA ATTCTGATA ACCCAAACAC
 20341 CTACGACTAC ATGAACAAGC GAGTGGTGGC TCCCAGGTTA GTGGACTGCT ACATTAACCT
 20401 TGGAGCACGC TGTTCCCTG ACTATATGGA CAACGTCAAC CCATTTAAC ACCACCGCAA
 20461 TGCTGCCCTG CGCTACCAGT CAATGTTGCT GGGCAATGGT CGCTATGTGC CCTTCCACAT
 20521 CCAGGTGCCCT CAGAAGTTCT TTGCCCCATAA AAACCTCCCTT CTCTGCCGG GCTCATACAC
 20581 CTACGAGTGG AACTTCAGGA AGGATGTTAA CATGGTTCTG CAGAGCTCCC TAGGAAATGAA
 20641 CCTAAGGGTT GACGGAGGCCA GCATTAAGTT TGATAGCATT TGCCCTTACG CCACCTTCTT
 20701 CCCCATGGCC CACAAACACCG CCTCCACGCT TGAGGCCATG CTTAGAAACG ACACCAACGA
 20761 CCAGTCTTT AACGACTATC TCTCCGGCG CAACATGTC TACCCATAC CCGCCAACGC
 20821 TACCAACGTG CCCATATCCA TCCCCCTCCCG CAACTGGCG GCTTTCCCG GCTGGGCCCTT
 20881 CACGCCCTT AAGACTAAGG AAACCCCATC ACTGGGCTCG GGCTACGACC CTTATTACAC
 20941 CTACTCTGGC TCTATACCT ACCTAGATGG AACCTTTAC CTCAACCACA CCTTTAAGAA
 21001 GGTGGCCATT ACCTTTGACT CTTCTGTAG CTGGCCTGGC AATGACGGCC TGCTTACCCC
 21061 CAACGAGTTT GAAAATTAAGC GTCAGTTGA CGGGGAGGGT TACAACGTTG CCCAGTGTAA
 21121 CATGACCAAA GACTGGTCC TGGTACAAAT GCTAGCTAAC TACAACATTG GCTACCAGGG
 21181 CTTCTATATC CCAGAGAGCT ACAAGGACCG CATGTACTCC TTCTTAGAA ACTTCCAGCC
 21241 CATGAGCCGT CAGGTGGTGG ATGATACTAA ATACAAGGAC TACCAACAGG TGGGCATCCT
 21301 ACACCAACAC AACAACTCTG GATTGTTGG CTACCTTGCC CCCACCATGC GCGAAGGACA
 21361 GGCCTACCC GCTAACTTCC CCTATCCGCT TATAGGCAAG ACCGAGTTG ACAGCATTAC
 21421 CCAGAAAAAG TTCTTTGCG ATCGCACCT TTGGCGCATC CCATTCTCCA GTAACTTTAT
 21481 GTCCATGGGC GCACTCACAG ACCTGGCCA AAACCTTCTC TACGCCAAT CCGCCACGC
 21541 GCTAGACATG ACTTTTGAGG TGGATCCAT GGACGAGGCC ACCCTTCTT ATGTTTGTT
 21601 TGAAGTCTTT GACGTGGTCC GTGTGCACCG GCGCACCG GCGTCATCG AAACCGTGTAA
 21661 CCTGGCACG CCCTCTCGG CGCGAACCG CACAACATAA AGAAGCAAGC AACATCAACAA
 21721 ACAGCTGCCG CCATGGGCTC CAGTGAGCAG GAACTGAAAG CCATTGTCAA AGATTTGGT
 21781 TGTGGCCAT ATTTTTGGG CACCTATGAC AAGCGTTTC CAGGCTTGT TTCTCCACAC
 21841 AAGCTGCCCT GCGCCATAGT CAATACGGCC GGTCGCGAGA CTGGGGCGT ACACTGGATG
 21901 GCCTTGCCT GGAACCCCGCA CTCAAAAACA TGCTACCTCT TTGAGCCCTT TGGCTTTCT
 21961 GACCAGCGAC TCAAGCAGGT TTACCAAGTT GAGTACGAGT CACTCCTGCG CCGTAGCGCC
 22021 ATTGCTCTT CCCCCGACCG CTGTATAACG CTGGAAAAGT CCACCCAAAG CGTACAGGGG
 22081 CCCAACTCGG CGGCCCTGTGG ACTATTCTGC TGCATGTTTC TCCACGCCCT TGCCAACCTGG
 22141 CCCCAAACTC CCATGGATCA CAACCCCAAC ATGAACCTTA TTACCGGGGT ACCCAAACCTCC
 22201 ATGCTCAACA GTCCCCAGGT ACAGCCCACC CTGCGTCGCA ACCAGGAACA GCTCTACAGC
 22261 TTCCTGGAGC GCCACTCGCC CTACTCCGC AGCCACAGTG CGCAGATTAG GAGGCCACT
 22321 TCTTTTGTC ACTTGAAAAA CATGTAAAAA TAATGTACTA GAGACACTTT CAATAAAGGC
 22381 AAATGTTTTT ATTTGTACAC TCTCGGGTGA TTATTTACCC CCACCCCTGC CGTCTCGGCC
 22441 GTTTAAAAAT CAAAGGGGTT CTGCCGCGCA TCGCTATGCG CCACTGGCAG GGACACGTTG
 22501 CGATACTGGT GTTTAGTGTCT CCACTAAAC TCAGGCACAA CCATCCGGG CAGCTCGGTG
 22561 AAGTTTCAC TCCACAGGCT GCGCACCATC ACCAACGCGT TTAGCAGGTC GGGGCCGAT
 22621 ATCTGAACT CGCAGTTGGG GCCTCCGCC TGCGCGCG AGTTGCGATA CACAGGGTTG
 22681 CAGCACTGGA ACACATATCAG CGCCGGGTGG TGCACTGCTGG CCAGCACGCT CTTGTCGGAG
 22741 ATCAGATCCG CGTCCAGGTC CTCCCGTTG CTCAGGGCGA ACGGAGTCAA CTTTGGTAGC
 22801 TGCCCTCCCA AAAAGGGCGC GTGCCCGAGG TTTGAGTTGC ACTCGCACCG TAGTGGCATC
 22861 AAAAGGTGAC CGTCCCCGGT CTGGCGTTA GGATACAGCG CCTGCATAAA AGCCTTGATC
 22921 TGCTTAAAG CCACCTGAGC CTTTGCCT TCAGAGAAGA ACATGCCGCA AGACTGCGC
 22981 GAAAACGTAT TGGCCGGACA GGCGCGCTCG TGCACTGCGAC ACCTTGCGTC GGTGTTGGAG
 23041 ATCTGCACCA CATTCTGGCC CCACCGGTTT TTCACGATCT TGGCCTTGCT AGACTGCTCC

64/92

23101 TTCAGCGCGC GCTGCCCGTT TTTCGCTCGTC ACATCCATT CAATCACGTG CTCCTTATTT
 23161 ATCATAATGC TTCCGTGTAG ACACCTAACG TCGCCTCGA TCTCAGCGCA GCGGTGCAGC
 23221 CACAACCGCG AGCCCCTGGG CTCGTATGC TTGTAGGTCA CCTCTGCAAA CGACTGCAGG
 23281 TACGCCCTGCA GGAATCGCCC CATCATCGTC ACAAAAGGTCT TGTTGCTGGT GAAGGTGCAGC
 23341 TGCAACCCGC GGTGCTCCTC GTTCAAGCCAG GTCTTGCATA CGGCCGCCAG AGCTTCCACT
 23401 TGGTCAGGCA GTAGTTGAA GTTCAAGCTT AGATCGTTAT CCACGTGGTA CTTGTCCATC
 23461 AGCGCGCGCG CAGCCTCCAT GCCCCTCTCC CACGCAGACA CGATCAGGCAC ACTCAGCGGG
 23521 TTCATCACCG TAATTTCACT TTCCGCTCG CTGGGCTCTT CCTCTTCCCT TTGCGTCCGC
 23581 ATACCACCGC CCACTGGGTC GTCTTCATTC AGCCGCCGCA CTGTGCGCTT ACCTCCCTTG
 23641 CCATGCTGAA TTAGCACCGG TGGGTTGCTG AAACCCACCA TTTGTAGCGC CACATCTTCT
 23701 CTTTCTTCCT CGCTGTCCAC GATTACCTCT GGTGATGGCG GGCCTCGGG CTTGGGAGAA
 23761 GGGCGCTTCT TTTTCTTCCTT GGGCGCAATG GCCAAATCCG CGCCCGAGGT CGATGGCCGC
 23821 GGGCTGGGTG TCGCGGGCAC CAGCGCGTCT TGTGATGAGT CTTCTCGTC CTCGGACTCG
 23881 ATACGCCGCC TCATCCGCTT TTTTGGGGC GCCCCGGGAG GCGGCGGGCGA CGGGGACCGGG
 23941 GACGACACGT CCTCCATGGT TGGGGGACGT CGCGCCGCAC CGCGTCCGCG CTCGGGGGTG
 24001 GTTTCGCGCT GCTCCCTCTC CCGACTGGCC ATTTCCTCT CCTATAGGCA GAAAAGATC
 24061 ATGGAGTCAG TCGAGAAAGAA GGACAGCTA ACCGGCCCTC CTGAGTTCCG CACCACCGCC
 24121 TCCACCGATG CCGCCAACGC GCCTTACCAAC TTCCCGCTCG AGGCACCCCCC GCTTGAGGAG
 24181 GAGGAAGTGA TTATCGAGCA GGACCCAGGT TTTGTAAGCG AAGACGACGA GGACCGCTCA
 24241 GTACCAACAG AGGATAAAAAA GCAAGACCG GACAACGCAG AGGCAAACGA GGAACAAGTC
 24301 GGGCGGGGGG ACGAAAGGCA TGGCGACTAC CTAGATGTGG GAGACGACGT GCTGTTGAAG
 24361 CATCTGCAGC GCCAGTGCAC CATTATCTGC GACCGCTTGC AAGACGCGAG CGATGTGCC
 24421 CTCGCCATAG CGGATGTCAAG CTTGCTTAC GAAAGCCACC TATTCTCACCG GCGCGTACCC
 24481 CCCAAACGCC AACAAAAACGG CACATGCGAG CCCAACCCGC GCCTCAACTT CTACCCCGTA
 24541 TTTGCCGTGC CAGAGGTGCT TGCCACCTAT CACATCTTTT TCCAAAACGT CAAGATAACCC
 24601 CTATCTGCC GTGCCAACCG CAGCCGAGCG GACAAGCAGC TGGCCTTGCG GCAGGGCGCT
 24661 GTCATACCTG ATATCGCCCTC GCTCAACGAA GTGCCAAAGA TCTTTGAGGG TCTTGACGC
 24721 GACGAGAACG CGCGGGCAAA CGCTCTGCAA CAGGAAAAAC GCGAAAATGA AAGTCACTCT
 24781 GGAGTGTGTTG TGGAACCTCGA GGGTGACAAAC CGCCGCCCTAG CCGTACTAAA ACGCAGCATH
 24841 GAGGTCAACCC ACTTTGCCA CCCGGCACTT AACCTACCCCC CCAAGGTCA GAGCACAGTC
 24901 ATGAGTGAGC TGATCGTGC CGGTGCGCAG CCCCTGGAGA GGGATGCAA TTTGCAAGAA
 24961 CAAACAGAGG AGGGCCTACC CGCAGTTGGC GACGAGCAGC TAGCGCGCTG GCTCAAACG
 25021 CGCGAGCCCTG CGGACTTGGG GGAGCGACGC AAACATAATGA TGGCCGCAGT GCTGTTACC
 25081 GTGGAGCTTG AGTGCATGCA GCGGTTCTTT GCTGACCCCG AGATGCAGCG CAAGCTAGAG
 25141 GAAACATTGC ACTACACCTT TCGACAGGGC TACGTACGCC AGGCCTGCAA GATCTCAAAC
 25201 GTGGAGCTCT GCAACCTGGT CTCCCTACCTT GGAATTTCG ACGAAAACCG CCTTGGGCAA
 25261 AACGTGCTTC ATTCCACGCT CAAGGGCGAG CGCGCCCGCG ACTACGTCCG CGACTGCGTT
 25321 TACTTATTTC TATGCTACAC CTGGCAGACG GCCATGGCG TTTGGCAGCA GTGCTTGGAG
 25381 GAGTGCAACC TCAAGGAGCT GCAGAAACTG CTAAAGCAA ACTTGAAGGA CCTATGGACG
 25441 GCCTTCAACG AGCGCTCCGT GGCGCGCAGC CTGGCGGACA TCATTTCCC CGAACGCCCTG
 25501 CTTAAAACCC TGCAACAGGG TCTGCCAGAC TTCACCAAGTC AAAGCATGTT GCAGAACTTT
 25561 AGGAACCTTA TCTTAGAGCG CTCAGGAATC TTGCCCGCCA CCTGCTGTG ACTTCCTAGC
 25621 GACTTTGTGC CCATTAAGTA CGCGGAATGC CCTCCGCCGC TTTGGGGCCA CTGCTACCTT
 25681 CTGCAGCTAG CCAACTACCT TGCCTACCAAC TCTGACATAA TGGAAAGACGT GAGCGGTGAC
 25741 GGTCTACTGG AGTGTACTG TCGCTGCAAC CTATGCACCC CGCACCGCTC CCTGGTTTGC
 25801 AATTGCAGC TGCTTAACGA AAGTCAAATT ATCGGTACCT TTGAGCTGCA GGGTCCCTCG
 25861 CCTGACGAAA AGTCCCGGGC TCCGGGGTTG AAACTCACTC CGGGGCTGTG GACGTCGGCT
 25921 TACCTCGCA AATTTGTACC TGAGGACTAC CACGCCACG AGATTAGGTT CTACGAAGAC
 25981 CAATCCCGCC CGCCAAATGC GGAGCTTACG GCCTGCGTCA TTACCCAGGG CCACATTCTT
 26041 GGCCAATTGC AAGCCATCAA CAAAGCCGC CAAGAGTTTC TGCTACGAA GGGACGGGGG
 26101 GTTTACTTGG ACCCCCCAGTC CGGCGAGGAG CTCAACCCAA TCCCCCCGCC GCCGCAGCCCC
 26161 TATCAGCAGC AGCGCGGGC CTTGCTTCC CAGGATGGCA CCCAAAAAGA AGCTGCAGCT
 26221 GCGGCCGCCA CCCACGGACG AGGAGGAATA CTGGGACAGT CAGGCAGAGG AGGTTTTGGA
 26281 CGAGGAGGAG GAGGACATGA TGGAAGACTG GGAGAGCCTA GACGAGGAAG CTTCCGAGGT
 26341 CGAAGAGGTG TCAGACGAAA CACCGTCACC CTGGTGCACG TTCCCTCGC CGGCCCGCCCA

FIG. 8H

65/92

26401 GAAATCGGCA ACCGGTTCCA GCATGGCTAC AACCTCCGCT CCTCAGGCAC CGCCGGCACT
 26461 GCCCCGTCGC CGACCCAACC GTAGATGGGA CACCACTGGA ACCAGGGCCG GTAAGTCCAA
 26521 GCAGCCGCCG CGGTTAGCCC AAGAGCAACA ACAGCGCCAA GGCTACCGCT CATGGCGCGG
 26581 GCACAAGAAC GCCATAGTTG CTTGCTTGCA AGACTGTGGG GGCAACATCT CCTTCGCCCC
 26641 CCGCTTCTT CTCTACCATC ACGGCGTGGC CTTCCCCCGT AACATCCTGC ATTACTACCG
 26701 TCATCTCTAC AGCCCCATACT GCACCGCGG CAGCGGCAGC GGCAGCAACA GCAGCGGCCA
 26761 CACAGAAGCA AAGGCAGCCG GATAGCAAGA CTCTGACAAA GCCCAAGAAA TCCACAGCGG
 26821 CGGAGCAGC AGGAGGAGGA GCGCTGCCTC TGGCGCCCAA CGAACCCGTA TCGACCCGCG
 26881 AGCTTAGAAA CAGGATTTT CCCACTCTGT ATGCTATATT TCAACAGAGC AGGGGCCAAG
 26941 ACAAAAGAGCT GAAAATAAAA AACAGGTCTC TGGCATCCCT CACCCGCAGC TGCCCTGTATC
 27001 ACAAAAGCGA AGATCAGCTT CGGCGCACGC TGGAAAGACGC GGAGGCTCTC TTCAGTAAAT
 27061 ACTGCGCGCT GACTCTTAAG GACTAGTTT GCGCCCTTTC TCAAATTAA GCGCGAAAAC
 27121 TACGTCTAC CCAGCGGCCA CACCCGGCGC CAGCACCTGT CGTCAGCGCC ATTATGAGCA
 27181 AGGAAATTCC CACGCCCTAC ATGTTGGAGTT ACCAGCCACA AATGGGACTT GCGGCTGGAG
 27241 CTGCCAAGA CTACTCAACC CGAATAAACT ACATGAGCGC GGGACCCAC ATGATATCCC
 27301 GGGTCAACCGG AATCCGCGCC CACCGAAACC GAATTCTCTT GGAACAGGCG GCTATTACCA
 27361 CCACACCTCG TAATAACCTT AATCCCCGTA GTGGCGCCG TGCCCTGGTG TACCAAGAAA
 27421 GTCCCGCTCC CACCACTGTG GTACTTCCC GAGACGCCA GGGCGAAGTT CAGATGACTA
 27481 ACTCAGGGGC GCAAGCTTGCG GGGCGCTTTC GTCAAGGGT GCGGTCGCC GGGCAGGGTA
 27541 TAACTCACCT GACAATCAGA GGGCGAGGTA TTCAGCTCAA CGACGAGTCG GTGAGCTCCT
 27601 CGCTTGGCTC CGCTCCGGAC GGGACATTTC AGATGGCGG CGGGCGCGT CCTTCATTCA
 27661 CGCCTCGTCA GCGAATCTTA ACTCTGCAGA CCTCGTCCTC TGAGCCGCGC TCTGGAGGCA
 27721 TTGGAACCTCT GCAATTATTATT GAGGAGTTTG TGCCATCGGT CTACTTTAAC CCCTTCTCGG
 27781 GACCTCCCGG CCACTATCCG GATCAATTAA TTCTTAACATT TGACGCGTA AAGGACTCGG
 27841 CGGACGGCTA CGACTGAATG TTAAGTGGAG AGGCAGAGCA ACTGCGCTG AAACACCTGG
 27901 TCCACTGTG CGGCCACAAG TGCTTGCCTC GCGACTCCGG TGAGTTTTGC TACTTTGAAT
 27961 TGCCCCAGGA TCATATCGAG GGGCCGGCGC ACGGCCTCCG GCTTACCGCC CAGGGAGAGC
 28021 TTGCCCCGTAG CCTGATTCTGG GAGTTTACCC AGGGCCCCCT GCTAGTTGAG CGGGACAGGG
 28081 GACCCCTGTGT TCTCACTGTG ATTTCGAACCT GTCTAACCT TGGATTACAT CAAGATCTTT
 28141 GTGCCATCT CTGTGCTGAG TATAATAAAAT ACAGAAATTAA AAATATACTG GGGCTCCTAT
 28201 CGCCATCCTG TAAACGCCAC CGTCCTCACC CGCCCAAGCA AACCAAGGCG AACCTTACCT
 28261 GGTACTTTTA ACATCTCTCC CTCTGTGATT TACAACAGTT TCAACCCAGA CGGAGTGAGT
 28321 CTACGAGAGA ACTCTCCGA GTCAGCTAC TCCATCAGAA AAAACACCCAC CCTCCTTACC
 28381 TGCCGGAAAC GTACGAGTGC GTCACGGCC GTCAGCTC ACCTACCGCC TGACCGTAAA
 28441 CCAGACTTTT TCCGGACAGA CCTCAATAAC TCTGTTTACCA AGAACAGGAG GTGAGCTTAG
 28501 AAAACCTTA GGGTATTAGG CCAAAGGCGC AGCTACTGTG GGGTTTATGA ACAATTCAAG
 28561 CAACTCTACG GGCTATTCTA ATTCAAGTTT CTCTAGAAC GGGGTTGGGG TTATCTCTG
 28621 TCTTGTGATT CTCTTATTTC TTATACAAAC GCTCTCTGC CTAAGGCTCG CGCCCTGCTG
 28681 TGTGCACATT TGCATTATTGT GTCAGCTTT TAAACGCTGG GGTGCCACC CAAGATGATT
 28741 AGGTACATAA TCTCTAGTTT ACTCACCCTT GCGTCAGCCC ACGGTACAC CCAAAAGGTG
 28801 GATTTTAAGG AGCCAGCTG TAATGTTACA TTGGCAGCTG AAGCTAATGA GTGCACCACT
 28861 CTTATAAAAT GCACCCACAGA ACATGAAAAG CTGCTTATTG GCCACAAAAA CAAAATTGGC
 28921 AAGTATGCTG TTTATGCTAT TTGGCAGCCA GGTGACACTA CAGAGTATAA TGTTACAGTT
 28981 TTCCAGGGTA AAAGTCATAA AACTTTTATG TATACTTTTC CATTCTATGA AATGTGCGAC
 29041 ATTACCATGT ACATGAGCAA ACAGTATAAG TTGTGGCCCC CACAAAATTG TGTGAAAAC
 29101 ACTGGCACTT TCTGCTGCAC TGCTATGCTA ATTACAGTGC TCGCTTTGGT CTGTACCCCA
 29161 CTCTATATTA AATACAAAAG CAGACGCAGC TTTATTGAGG AAAAGAAAAT GCCTTAATT
 29221 ACTAAGTTAC AAAGCTAATG TCACCACTAA CTGCTTACT CGCTGCTTG AAAACAAATT
 29281 CAAAAAGTTA GCATTATAAT TAGAATAGGA TTTAAACCCC CGGGTCATTT CCTGCTCAAT
 29341 ACCATCCCC TGAACAATTG ACTCTATGTG GGATATGCTC CAGCGCTACA ACCTTGAAGT
 29401 CAGGCTTCCT GGATGTCAGC ATCTGACTTT GGCCAGCACC TGTCGGGGG ATTTGTTCCA
 29461 GTCCAACCTAC AGCGACCCAC CCTAACAGAG ATGACCAACA CAACCAACGC GGCCGCCGCT
 29521 ACCGGACTTA CATCTACCAAC AAATACACCC CAAGTTCTG CCTTGTCAA TAACTGGGAT
 29581 AACTTGGCA TGTGGTGGTT CTCCATAGCG CTTATGTTTG TATGCCTTAT TATTATGTGG
 29641 CTCATCTGCT GCCTAAAGCG CAAACGCGCC CGACCACCA TCTATAGTCC CATCATTGTG

FIG. 81

66/92

29701 CTACACCCAA ACAATGATGG AATCCATAGA TTGGACGGAC TGAAACACAT GTTCTTTCT
 29761 CTTACAGTAT GATTAATGA GACATGATTG CTCGAGTTT TATATTACTG ACCCTTGTG
 29821 CGCTTTTTG TGCGTGCCTC ACATTGGCTG CGGTTTCTCA CATCGAAGTA GACTGCATTC
 29881 CAGCCTTCAC AGCTTATTG CTTTACGGAT TTGTCACCCT CACGCTCATC TGCAAGCTCA
 29941 TCACTGTGGT CATGCCCTT ATCCAGTGCA TTGACTGGGT CTGTGTGCGC TTTGCATATC
 30001 TCAGACACCA TCCCCAGTAC AGGGACAGGA CTATAGCTGA GCTTCTTAGA ATTCTTTAAT
 30061 TATGAAATT ACTGTGACTT TTCTGCTGAT TATTGCAAC CTATCTGCGT TTTGTTCCCC
 30121 GACCTCCAAG CCTCAAAGAC ATATATCATG CAGATTCACT CGTATATGGA ATATTCCAAG
 30181 TTGCTACAAAT GAAAAAAAGCG ATCTTTCGGA AGCCTGGTTA TATGCAATCA TCTCTGTTAT
 30241 GGTGTTCTGC AGTACCATCT TAGCCCTAGC TATATATCCC TACCTTGACA TTGGCTGGAA
 30301 ACGAATAGAT GCCATGAACC ACCCAACTT CCCCCGCGCC GCTATGCTTC CACTGCAACA
 30361 AGTTGTGCC GGCGGCTTTG TCCCAGCCAA TCAGGCTCGC CCCACTTCTC CCACCCCCAC
 30421 TGAAATCAGC TACTTTAAC TAAACAGGAGG AGATGACTGA CACCCCTAGAT CTAGAAATGG
 30481 ACGGAATTAT TACAGAGCAG CGCCCTGCTAG AAAGACGCAG GGCAGCGGCC GAGCAACAGC
 30541 GCATGAATCA AGAGCTCCAA GACATGGTTA ACTTGCACCA GTGCAAAAGG GGTATCTTT
 30601 GTCTGGTAAA GCAGGCCAAAG GTCACCTACG ACAGTAATAC CACCGGACAC CGCCTTAGCT
 30661 ACAAGTGCC AACCAAGCGT CAGAAATTGG TGGTCATGGT GGGAGAAAAG CCCATTACCA
 30721 TAACTCAGCA CTCGGTAGAA ACCGAAAGCT GCATTCACTC ACCTTGTCAA GGACCTGAGG
 30781 ATCTCTGCAC CCTTATTAAG ACCCTGTGCG GTCTCAAAGA TCTTATTCCTC TTTAACTAAT
 30841 AAAAAAAAT AATAAAGCAT CACTTACTTA AAATCAGTTA GCAAATTCT GTCCAGTTA
 30901 TTCAGCAGCA CCTCTTGCC CTCCTCCAG CTCTGGTATT GCAGCTTCCT CCTGGCTGCA
 30961 AACTTCTCC ACAATCTAAA TGGAATGTCA GTTCCCTCCT GTTCCTGTCC ATCCGCACCC
 31021 ACTATCTCA TGTGTTGCA GATGAAGCGC GCAAGACCGT CTGAAGATAAC CTTCAACCCC
 31081 GTGTATCCAT ATGACACCGA AACCGGTCTC CCAACTGTGC CTTTCTCTTAC TCCTCCCTT
 31141 GTATCCCCCA ATGGGTTTCA AGAGAGTCCC CCTGGGGTAC TCTCTTGCG CCTATCCGAA
 31201 CCTCTAGTTA CCTCCAATGG CATGCTTGCG CTCAAAATGG GCAACGGCCT CTCTCTGGAC
 31261 GAGGCCGGCA ACCTTACCTC CCAAAATGTA ACCACTGTGA GCCCACCTCT CAAAAAAACC
 31321 AAGTCAAACA TAAACCTGGA AATATCTGA CCCCTCACAG TTACCTCAGA AGCCCTAACT
 31381 GTGGCTGCGC CCGCACCTCT AATGGTCGG GGCACACAC TCACCATGCA ATCACAGGCC
 31441 CCGCTAACCG TGACAGACTC CAAACTTAGC ATTGCCACCC AAGGACCCCT CACAGTGTCA
 31501 GAAGGAAAGC TAGCCCTGCA AACATCAGGC CCCCTCACCA CCACCGATAG CAGTACCCCT
 31561 ACTATCACTG CCTCACCCCCC TCTAACTACT GCCACTGGTA GCTTGGGCAT TGACTGAAA
 31621 GAGCCCCATT ATACACAAAAA TGGAAACCTA GGACTAAAGT ACGGGGCTCC TTTGCATGTA
 31681 ACAGACGACC TAAACACTTT GACCGTAGCA ACTGGTCCAG GTGTGACTAT TAATAACT
 31741 TCCCTGCAAA CTAAGTTAC TGGAGCCTTG GTTGTGATT CACAAGGCAA TATGCAACTT
 31801 AATGTAGCAG GAGGACTAAG GATTGATTCT CAAACACAG GCCTTATACT TGATGTTAGT
 31861 TATCCGTTTG ATGCTAAAAA CCAACTAAAT CTAAGACTAG GACAGGGCCC TCTTTTATA
 31921 AACTCAGCCC ACAACTTGG AATTAACCTAC AACAAAGGCC TTTACTTGT TACAGCTTC
 31981 AACAAATTCCA AAAAGCTTGA GTTAAACCTA AGCACTGCCA AGGGGTTGAT GTTGACGCT
 32041 ACAGCCATAG CCATTAATGC AGGAGATGGG CTTGAATTG GTTCACCTAA TGCAACAAAC
 32101 ACAAAATCCCCC TCAAAACAAA AATTGGCCAT GGCTAGAAT TTGATTCAA CAAGGCTATG
 32161 GTTCTAAAC TAGGAACCTGG CCTTAGTTT GACAGCACAG GTGCCATTAC AGTAGGAAAC
 32221 AAAAATAATG ATAAGCTAAC TTTGTGGACC ACACCAGCTC CATCTCCTAA CTGTAGACTA
 32281 AATGCAGAGA AAAGATGCTAA ACTCACTTTG GTCTTAACAA AATGTGGCAG TCAAATACTT
 32341 GCTACAGTTT CAGTTTGGC TGTAAAGGC AGTTGGCTC CAATATCTGG AACAGTTCAA
 32401 AGTGTCTCAT TTATTATAAG ATTTGACGAA ATGGAGTGC TACTAAACAA TTCTCTCTG
 32461 GACCCAGAAT ATGGAAACTT TAGAAATGGA GATCTTACTG AAGGCACAGC CTATACAAAC
 32521 GCTGTGGAT TTATGCCCTAA CCTATCAGCT TATCCAAAAT CTCACGGTAA AACTGCCAA
 32581 AGTAACATTG TCAGTCAAGT TTACTTAAAC GGAGACAAAAA CTAAACCTGT AACACTAAC
 32641 ATTACACTAA ACGGTACACA GGAAACAGGA GACACAACTC CAAGTGCATA CTCTATGTCA
 32701 TTTTCATGGG ACTGGTCTGG CCACAACTAC ATTAATGAAA TATTGCCCAC ATCCTTTAC
 32761 ACTTTTCAT ACATTGCCCA AGAATAAAGA ATCGTTTGTG TTATGTTTCA ACGTGTTTAT
 32821 TTTTCATATTG CAGAAAATT CAAGTCATT TTCATTCAAGT AGTATAGCCC CACCACCA
 32881 TAGCTTATAC AGATCACCGT ACCTTAATCA AACTCACAGA ACCCTAGTAT TCAACCTGCC
 32941 ACCTCCCTCC CAACACACAG AGTACACAGT CCTTTCTCCC CGGCTGGCCT TAAAAGCAT

FIG. 8J

67/92

33001 CATAATCATGG GTAACAGAGACA TATTCTTAGG TGTTATATT CACACGGTTT CCTGTCGAGC
 33061 CAAACGCTCA TCAGTGATAT TAATAAACTC CCCGGGCAGC TCACTTAAGT TCATGTCGCT
 33121 GTCCAGCTGC TGAGCCACAG GCTGCTGTCC AACCTGCGGT TGCTTAACGG CGGGCGAAGG
 33181 AGAAGTCCAC GCCTACATGG GGGTAGAGTC ATAATCGTGC ATCAGGATAG GGCGGTGGTG
 33241 CTGCAGCAGC GCGCGAATAA ACTGCTGCCG CGCCGCTCC GTCCTGCAGG AATACAACAT
 33301 GGCAGTGGTC TCCTCAGCGA TGATTCCGAC CGCCCGCAGC ATAAGGCGCC TTGTCCTCCG
 33361 GGCACAGCAG CGCACCCCTGA TCTCACTTAA ATCAGCACAG TAACTGCAGC ACAGCACCAC
 33421 AATATTGTC AAAATCCCAC AGTGCAAGGC GCTGTATCCA AAGCTCATGG CGGGGACCAC
 33481 AGAACCCACG TGGCCATCAT ACCACAAGCG CAGGTAGATT AAGTGGCGAC CCCTCATAAA
 33541 CACGCTGGAC ATAAACATTA CCTCTTTTGG CATGTTGTAATTCACCCACCT CCCGGTACCA
 33601 TATAAACCTC TGATTAAACCA TGGGCCATC CACCACCATC CTAAACCCAGC TGGCAAAAC
 33661 CTGCCCGCCG GCTATACACT GCAGGGAACC GGGACTGGAA CAATGACAGT GGAGAGCCCC
 33721 GGACTCGTAA CCATGGATCA TCATGCTCGT CATGATATCA ATGTTGGCAC AACACAGGCA
 33781 CACGTGCATA CACTCCCTCA GGATTACAAG CTCTCTCCGC GTTAAAGACCA TATCCCAGGG
 33841 ACAACCCAT TCCTGAATCA GCGTAAATCC CACACTGCAG GGAAGACCTC GCACGTAACT
 33901 CACGTGTGC ATTGTCAAAG TGTTACATTC GGGCAGCAGC GGATGATCCT CCAGTATGGT
 33961 AGCGGGGTT TCTGTCTCAA AAGGAGGTAG ACGATCCCTA CTGTAACGGAG TGGCCGAGA
 34021 CAACCGAGAT CGTGTGGTC GTAGTGTCAATGCCAAATGGA ACGCCGGACG TAGTCATATT
 34081 TCCCTGAAGCA AAACCAAGGTG CGGGCGTGAC AAACAGATCT GCGTCTCCGG TCTCGCCGCT
 34141 TAGATCGCTC TGTGTAGTAG TTGTAGTATA TCCACTCTCT CAAAGCATCC AGGGCCCCC
 34201 TGGCTTCGGG TTCTATGTAA ACTCCTTCAT GCCCGCGCTGC CCTGATAACA TCCACCACCG
 34261 CAGAATAAGC CACACCCAGC CAACCTACAC ATTCTGTTCTG CGAGTCACAC ACGGGAGGAG
 34321 CGGGAAAGAGC TGGAAGAACC ATGTTTTTTT TTTTATTCCA AAAGATTATC CAAACCTCA
 34381 AAATGAAGAT CTATTAAGTG AACGCGCTCC CCTCCGGTGG CGTGGTCAAA CTCTACAGCC
 34441 AAAGAACAGA TAATGGCATT TGTAAGATGT TGCAACATGG CTTCCAAAAG GCAAAACGGCC
 34501 CTCACGTCCA AGTGGACGTA AAGGCTAAAC CCTTCAGGGT GAATCTCCCTC TATAAACATT
 34561 CCAGCCTCTT CAACCATGCC CAAATAATT TCATCTCGCC ACCTTCTCAA TATATCTCTA
 34621 AGCAAAATCCC GAATATTAAG TCCGGCATT GTAAAAATCT GCTCCAGAGC GCCCTCCACC
 34681 TTCAGCCTCA AGCAGGAAAT CATGATTGCA AAAATTCAAGG TTCTCACAG ACCTGTATAA
 34741 GATTCAAAAG CGGAACATTAA CAAAAAAATAC CGCGATCCCG TAGGTCCCCCT CGCAGGGCCA
 34801 GCTGAACATA ATCGTGCAGG TCTGCACCGA CCAGCGCGGC CACTTCCCCG CCAGAACCT
 34861 TGACAAAAGA ACCCACACTG ATTATGACAC GCATACTCGG AGCTATGCTA ACCAGCGTAG
 34921 CCCCCGATGTA AGCTTTGTTG CATGGCGGC GATATAAAAT GCAAGGTGCT GCTCAAAAAA
 34981 TCAGGCAAAG CCTCGCGCAA AAAAGAAAGC ACATCGTAGT CATGCTCATG CAGATAAAGG
 35041 CAGGTAAAGCT CCGGAACCAC CACAGAAAAA GACACCATT TTCTCTCAA CATGCTGCG
 35101 GGTTCTGCA TAAACACAAA ATAAAATAAC AAAAAAAACAT TTAAACATTA GAAGCCTGTC
 35161 TTACACACAGG AAAAACAAACC CTTATAAGCA TAAGACGGAC TACGGCCATG CCGCGTGAC
 35221 CGTAAAAAAA CTGGTCACCG TGATTAAGA GCACCCACCGA CAGCTCCTG GTCATGTCCG
 35281 GAGTCATAAT GTAAGACTCG GTAAACACAT CAGGTTGATT CATCGTCAG TGCTAAAAAG
 35341 CGACCGAAAT AGCCCCGGGG AATACATACC CGCAGCGTA GAGACAAACAT TACAGCCCCC
 35401 ATAGGAGGTA TAACAAAATT AATAGGAGAG AAAAACACAT AAACACCTGA AAAACCTCC
 35461 TGCCTAGGCA AAATAGCACC CTCCCGCTCC AGAACAAACAT ACAGCGCTTC ACAGCGGCAG
 35521 CCTAACAGTC AGCCTTACCA GTAAAAAAGA AAACCTATTA AAAAAACACC ACTCGACACG
 35581 GCACCGAGCTC AATCAGTCAC AGTGTAAAAA AGGGCCAAGT GCAGAGCGAG TATATATAGG
 35641 ACTAAAAAAT GACGTAACGG TTAAAGTCCA CAAAAAAACAC CCAGAAAACC GCACCGGAAC
 35701 CTACGCCAG AAACGAAAGC CAAAAAAACCC ACAACTTCCT CAAATCGTCA CTTCCGTTT
 35761 CCCACGTTAC GTAACATTCCC ATTTTAAGAA AACTACAATT CCCAACACAT ACAAGTTACT
 35821 CGGCCCTAAA ACCTACGTCA CCCGCCCCGT TCCCACGCC CGCGCCACGT CACAAACTCC
 35881 ACCCCCCCTCAT TATCATATTG GCTTCACATCC AAAATAAGGT ATATTATTGTA TGATG

FIG. 8K

68/92

Structure of the Ad6 Genome

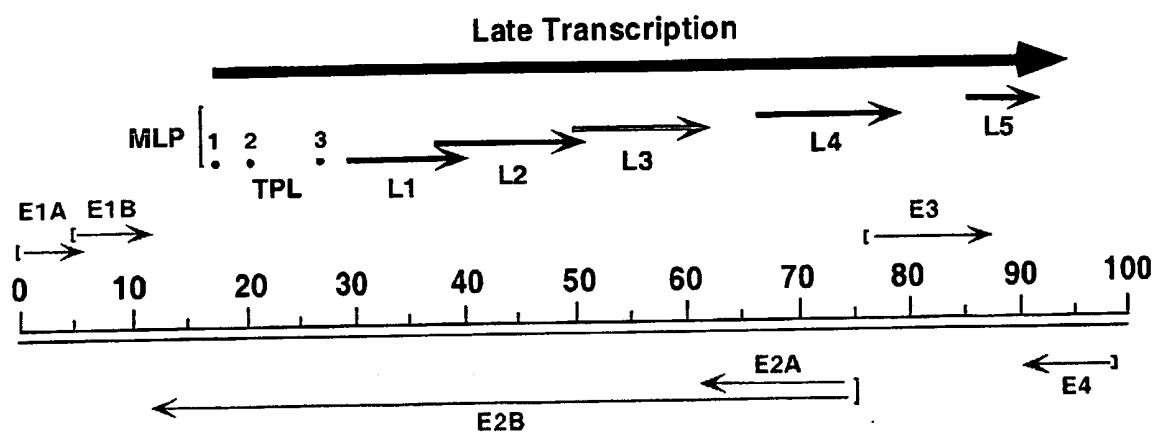


FIG. 9

69/92

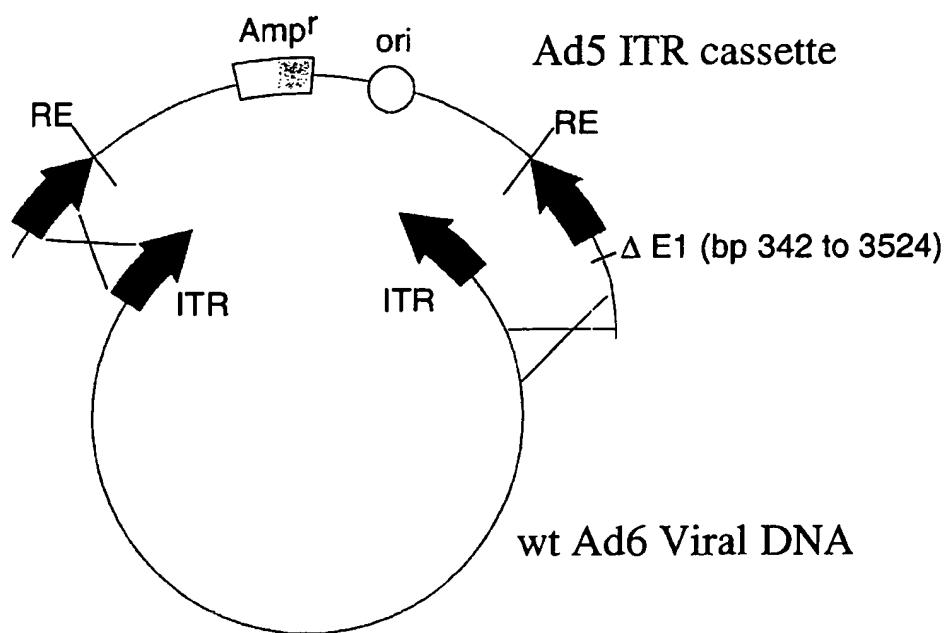


FIG. 10

70/92

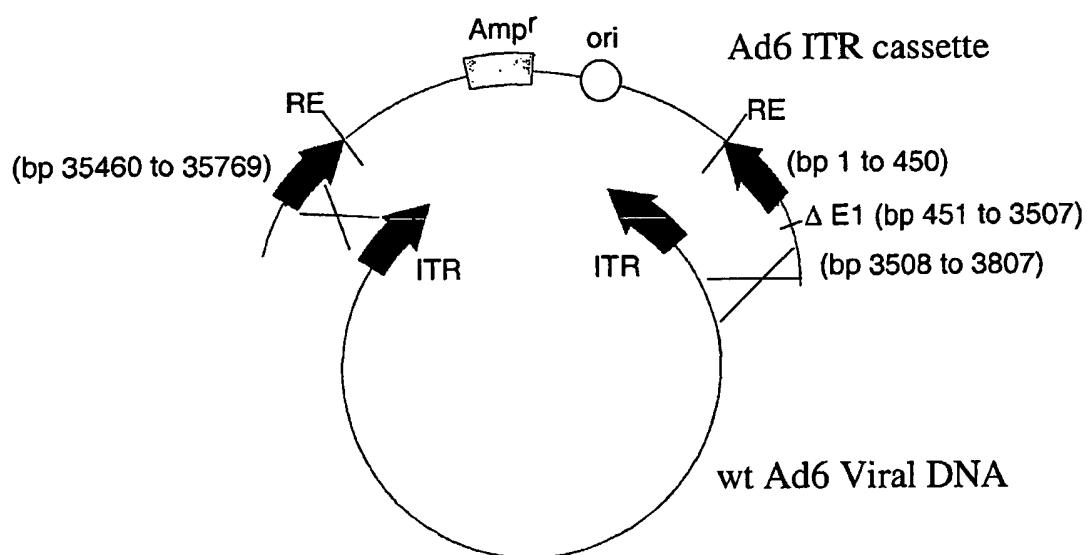
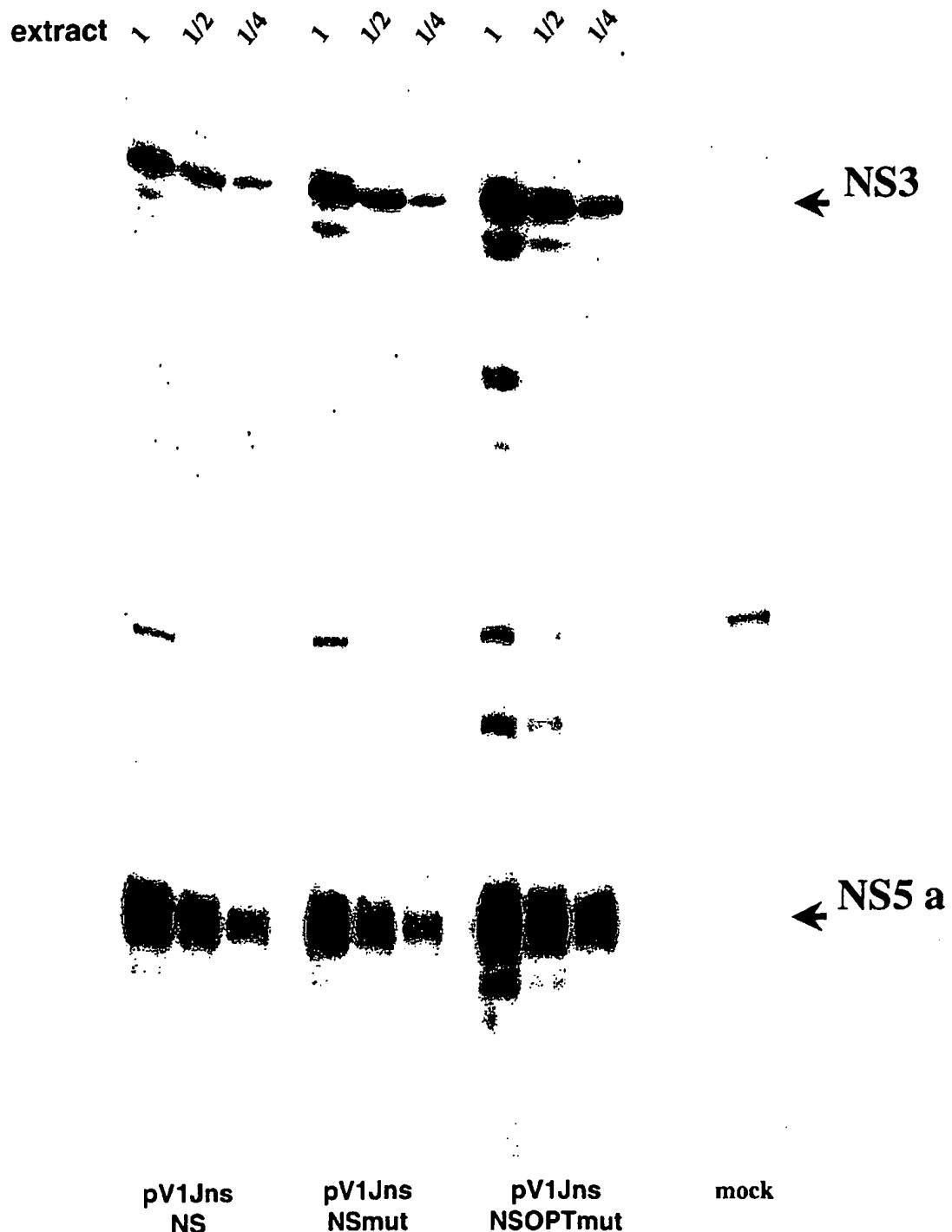


FIG. 11

71/92



Western blot on whole-cell extracts from 293 cells transfected with plasmid DNA expressing the different HCV NS cassettes. Mature NS3 and NSSA products were detected with specific antibodies.

FIG. 12

72/92

		Pep pool							
		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSO
pV1jns-NS	#31	41	135	19	44	25	17	137	8
	#32	121	783	77	144	13	22	604	4
	#33	8	32	3	11	6	6	43	3
	#34	16	139	13	47	31	25	151	2
	#35	21	101	40	32	21	20	75	1
	#36	18	26	24	25	5	7	29	6
	#37	19	73	15	39	8	20	49	2
	#38	133	575	74	345	75	63	515	5
	#39	40	183	10	85	14	9	148	2
	#40	66	465	29	111	15	16	189	0
Geomean		33	146	21	57	15	16	123	na

		Pep pool							
		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSO
pV1jns-NSmut	#41	39	293	58	187	5	4	248	1
	#42	21	220	46	107	26	10	189	4
	#43	76	134	12	78	8	6	144	2
	#44	30	45	20	52	4	8	40	4
	#45	36	100	17	56	4	6	116	3
	#46	67	172	16	138	8	9	145	3
	#47	34	131	28	38	9	5	118	1
	#48	55	316	43	107	9	7	277	5
	#49	6	131	5	25	4	1	91	0
	#50	13	93	11	11	5	1	76	1
Geomean		30	142	20	61	7	5	126	na

		Pep pool							
		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSO
V1jns-NSOPTmut	#51	53	409	34	84	11	25	271	4
	#52	140	660	65	276	23	36	377	2
	#53	58	553	48	105	23	18	564	1
	#54	50	105	35	134	10	16	80	2
	#55	14	80	11	35	4	7	91	6
	#56	14	342	30	101	23	14	207	1
	#57	63	325	66	239	17	24	123	1
	#58	75	542	66	168	127	93	191	0
	#59	65	468	40	124	18	23	344	4
	#60	27	142	48	16	7	8	77	0
Geomean		45	295	40	99	16	20	188	na

IFN γ ELIspot on splenocytes from C57black6 mice immunized with two injections of 25 μ g DNA/dose with GET of plasmid vectors expressing the different HCV NS cassettes. Data are expressed as SFC/10 6 PBMC.

FIG. 13A

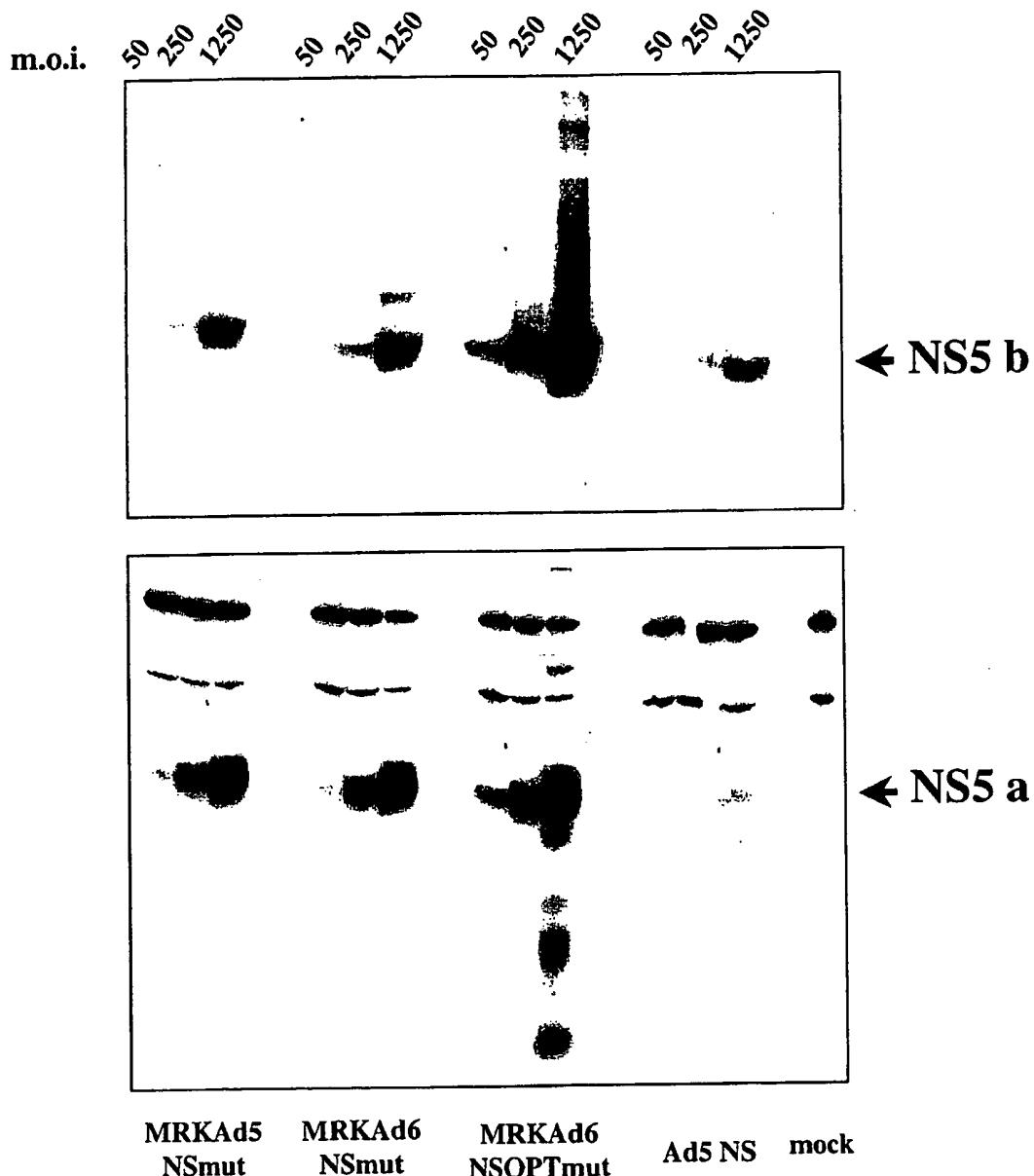
73/92

Pep pool								
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
pV1jns-NS	#51	219	699	634	486	487	264	34
	#52	67	302	347	167	111	87	9
	#53	59	460	400	246	244	136	26
	#54	139	817	685	236	547	223	24
	#55	96	904	542	277	256	337	17
	#56	225	603	686	156	350	240	56
	#57	44	288	211	148	100	141	4
	#58	37	262	221	53	58	62	3
	#59	131	975	928	159	305	284	14
	#60	93	475	464	77	206	113	12
geo mean		111	579	512	201	266	189	20
Pep pool								
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
pV1jns-NSmut	#61	72	840	515	219	278	249	19
	#62	294	1881	1266	365	434	411	63
	#63	73	415	422	103	141	99	41
	#64	66	824	486	175	162	144	18
	#66	24	313	168	53	47	42	5
	#67	15	230	253	94	25	39	2
	#68	53	354	252	89	101	86	15
	#69	271	895	909	518	322	285	74
	#70	417	1303	1186	468	557	267	34
	geo mean	143	784	606	232	230	180	30
Pep pool								
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
V1jns-NSOPTmut	#71	206	944	890	342	207	397	47
	#72	393	1655	1151	575	626	401	72
	#73	123	522	515	319	223	198	21
	#74	500	1414	1419	878	1035	1122	137
	#75	286	812	873	382	543	267	31
	#76	224	1143	942	218	420	281	22
	#77	95	643	630	169	385	218	15
	#78	401	1302	1068	538	608	623	12
	#79	108	1190	914	199	265	215	4
	#80	122	511	546	189	286	190	13
geo mean		209	941	854	331	406	329	24

IFN γ ELIspot on splenocytes from BalbC mice immunized with two injections of 50 μ g DNA/dose with GET of plasmid vectors expressing the different HCV NS cassettes. Data are expressed as SFC/10 6 PBMC.

FIG. 13B

74/92



Western blot on whole-cell extracts from HeLa cells infected at different multiplicity of infection (m.o.i.; indicated at the top) with Adenovectors expressing the different HCV NS cassettes. Mature NS5B and NS5A products were detected with specific antibodies.

FIG. 14

75/92

mouse	Pep pool							
	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)	DMSO	
Ad5-NS	#1	14	492	9	27	10	554	7
	#2	8	440	2	26	5	438	0
	#3	12	92	5	12	7	73	4
	#4	16	388	6	40	6	228	2
	#6	8	210	4	31	3	238	3
	#7	7	133	13	16	0	128	9
	#8	11	342	25	55	22	267	12
	#9	5	345	0	45	5	285	3
	#10	22	888	3	65	25	799	1
	Geomean	10	305	na	31	na	269	na
MRKAd5-NSmut	Pep pool							
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)	DMSO
	#11	14	1009	13	75	7	751	6
	#12	15	695	3	39	9	552	1
	#13	12	389	4	20	7	352	3
	#14	7	459	6	50	1	274	1
	#15	5	549	3	22	6	485	0
	#16	10	631	1	6	4	600	3
	#17	5	257	3	9	1	245	3
	#18	13	659	6	43	7	555	1
	#19	12	758	1	37	5	669	0
	#20	22	1380	5	163	8	1003	4
	Geomean	10	615	3	31	4	504	na
MRKAd6-NSmut	Pep pool							
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)	DMSO
	#21	6	584	5	27	4	491	2
	#22	6	231	3	12	3	235	0
	#23	8	482	1	18	1	511	0
	#24	14	1120	6	38	10	1004	5
	#25	1	311	3	9	0	382	1
	#26	29	903	3	60	5	751	5
	#27	35	1573	4	40	4	1277	4
	#28	7	406	5	15	1	443	3
	#29	4	461	3	12	3	515	3
	Geomean	8	567	3	21	na	554	na

IFN γ ELISPOT on splenocytes from C57black6 mice immunized with two injections of 10^9 vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 15

76/92

Pep pools	Ad5-NS 10^{10} vp/dose		
	96074	I34T	063Q
F (NS3p)	374	11	74
G (NS3h)	359	1070	1455
H (NS4)	376	30	64
I (NS5a)	240	40	63
L (NS5b)	226	29	121
M (NS5b)	511	23	35
DMSO	128	3	31

Pep pools	MRK Ad6-NSmut 10^{10} vp/dose		
	S207	035Q	057Q
F (NS3p)	363	382	150
G (NS3h)	180	316	119
H (NS4)	126	113	62
I (NS5a)	1780	688	114
L (NS5b)	447	111	81
M (NS5b)	153	38	16
DMSO	9	6	9

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with one injection of 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16A

77/92

Pep pools	MRK Ad5-NSmut 10^{10} vp/dose		
	S20I	075Q	137Q
F (NS3p)	928	69	254
G (NS3h)	317	436	98
H (NS4)	56	101	45
I (NS5a)	1530	1100	413
L (NS5b)	149	23	92
M (NS5b)	398	32	80
DMSO	29	6	29

Pep pools	MRK Ad6-NSOPTmut 10^{10} vp/dose		
	98D209	106Q	113Q
F (NS3p)	3110	263	404
G (NS3h)	2115	642	1008
H (NS4)	373	72	19
I (NS5a)	103	37	347
L (NS5b)	149	22	10
M (NS5b)	314	428	- 19
DMSO	0	1	3

IFNy ELISPOT on PBMC from Rhesus monkeys immunized with one injection of 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16B

78/92

Pep pools	Ad5-NS 10^{11} vp/dose			
	99C008	97N104	97X008	99C026
F (NS3p)	28	1026	579	889
G (NS3h)	1279	188	103	2453
H (NS4)	18	39	138	109
I (NS5a)	131	1068	172	141
L (NS5b)	78	144	103	32
M (NS5b)	24	68	47	84
DMSO	3	16	1	19

Pep pools	MRKAd6-NSmut 10^{11} vp/dose			
	98C047	97C055	93G	97X014
F (NS3p)	477	25	93	1022
G (NS3h)	959	398	81	1513
H (NS4)	36	14	99	53
I (NS5a)	171	45	1237	98
L (NS5b)	18	32	23	51
M (NS5b)	88	4	13	40
DMSO	8	3	1	5

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16C

79/92

Pep pools	MRKAd5-NSmut 10 ¹¹ vp/dose			
	99C059	99C060	97X009	96069
<i>F</i> (<i>NS3p</i>)	28	81	1308	1618
<i>G</i> (<i>NS3h</i>)	2600	161	1008	123
<i>H</i> (<i>NS4</i>)	31	74	101	40
<i>I</i> (<i>NS5a</i>)	181	99	69	96
<i>L</i> (<i>NS5b</i>)	24	31	40	20
<i>M</i> (<i>NS5b</i>)	11	58	38	164
<i>DMSO</i>	6	15	1	16

IFNy ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 10¹¹ vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/10⁶ PBMC.

FIG. 16D

80/92

Pep pools	MRK Ad5-NSmut 10 10 vp/dose		
	S201	075Q	137Q
pool F (NS3p)	881	1755	73
pool G (NS3h)	573		
pool H (NS4)		3541	
pool I (NS5a)	2094		39
pool L (NS5b)			
pool M (NS5b)	756		
DMSO	319	117	44

Pep pools	MRK Ad6-NSOPTmut 10 10 vp/dose		
	98D209	106Q	113Q
pool F (NS3p)	5073	84	952
pool G (NS3h)	2376	160	3325
pool H (NS4)	700		
pool I (NS5a)			1106
pool L (NS5b)			
pool M (NS5b)	530	706	
DMSO	43	47	28

Pep pools	MRK Ad6-NSmut 10 10 vp/dose		
	S207	035Q	057Q
pool F (NS3p)	118	480	
pool G (NS3h)		196	
pool H (NS4)			
pool I (NS5a)	3340	933	
pool L (NS5b)	118		
pool M (NS5b)			
DMSO	145	34	

IFN γ ICS on PBMC from Rhesus monkeys immunized with two injections at four weeks interval with 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as number of positive IFN γ /CD3/CD8 per 10^6 lymphocytes.

FIG. 17A

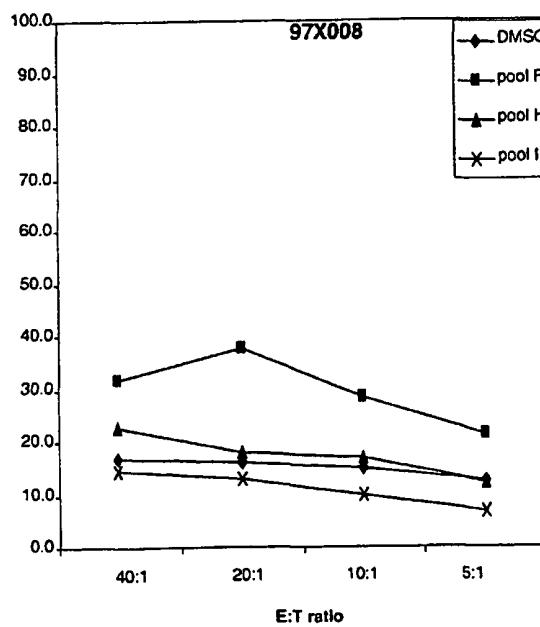
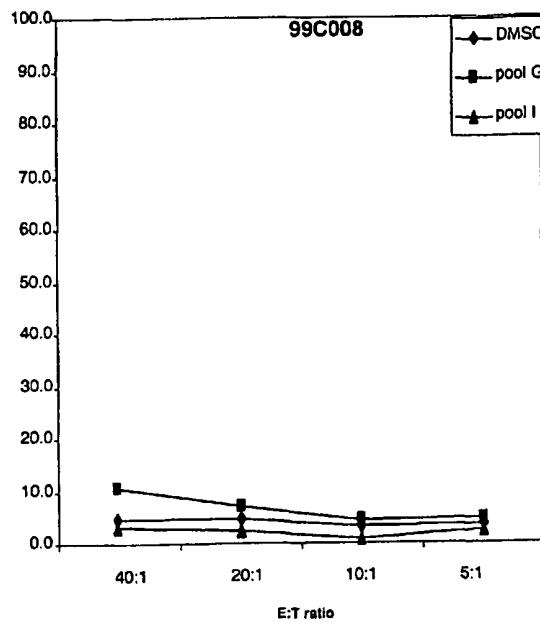
81/92

		Ad5-NS 10 ¹¹ vp/dose			
Pep pools		99C008	97N104	97X008	99C026
<i>F</i> (<i>NS3p</i>)			1703	1136	615
<i>G</i> (<i>NS3h</i>)		3153			2787
<i>H</i> (<i>NS4</i>)					
<i>I</i> (<i>NS5a</i>)			2233		
<i>L</i> (<i>NS5b</i>)					
<i>M</i> (<i>NS5b</i>)					
<i>DMSO</i>		125	98	130	0
		MRKAd6-NSmut 10 ¹¹ vp/dose			
Pep pools		98C047	97C055	93G	97X014
<i>F</i> (<i>NS3p</i>)		1024			948
<i>G</i> (<i>NS3h</i>)		3246	353		1074
<i>H</i> (<i>NS4</i>)				316	
<i>I</i> (<i>NS5a</i>)				6224	
<i>L</i> (<i>NS5b</i>)					
<i>M</i> (<i>NS5b</i>)					
<i>DMSO</i>		49	23	37	93
		MRKAd5-NSmut 10 ¹¹ vp/dose			
Pep pools		99C059	99C060	97X009	96069
<i>F</i> (<i>NS3p</i>)				2266	5053
<i>G</i> (<i>NS3h</i>)		2434	316	1018	
<i>H</i> (<i>NS4</i>)					
<i>I</i> (<i>NS5a</i>)					
<i>L</i> (<i>NS5b</i>)					205
<i>M</i> (<i>NS5b</i>)					
<i>DMSO</i>		13	110	119	15

IFNy ICS on PBMC from Rhesus monkeys immunized with two injections at four weeks interval with 10¹¹ vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as number of positive IFNy/CD3/CD8 per 10⁶ lymphocytes.

FIG. 17B

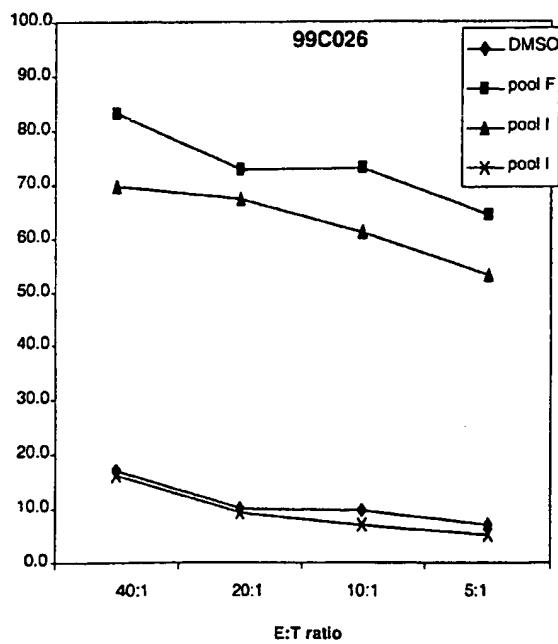
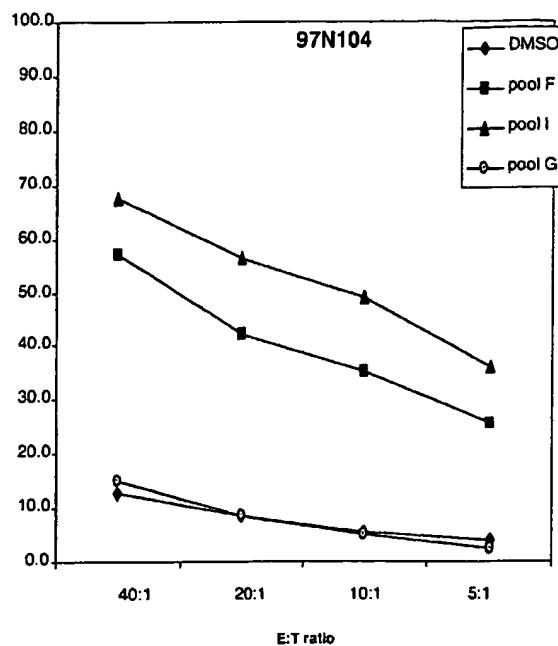
82/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Ad5-NS.

FIG. 18A

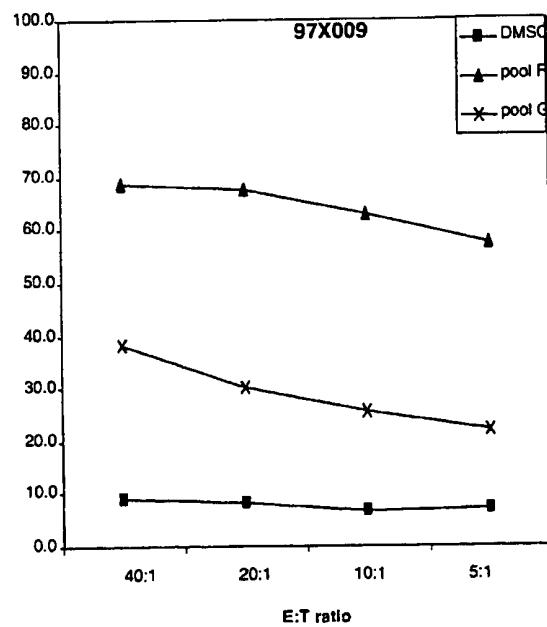
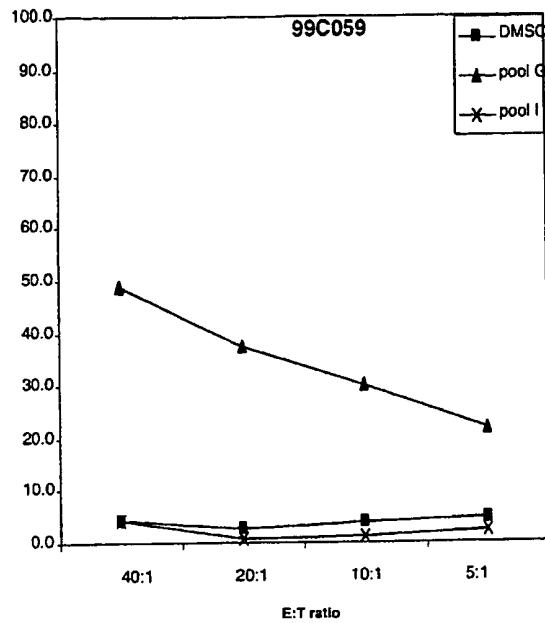
83/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Ad5-NS.

FIG. 18B

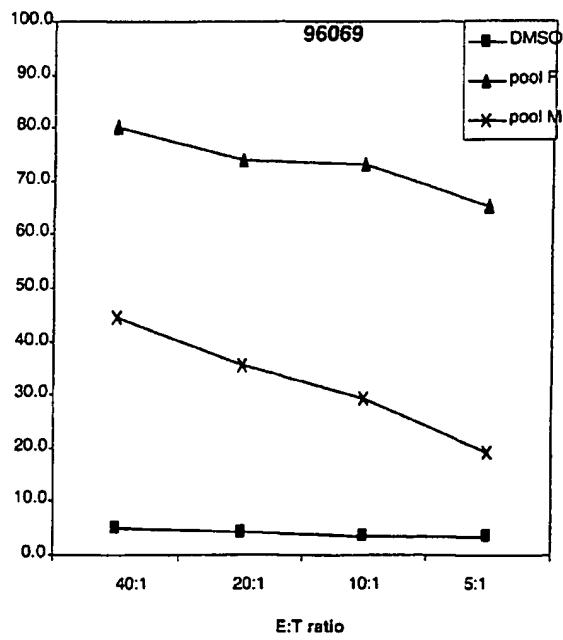
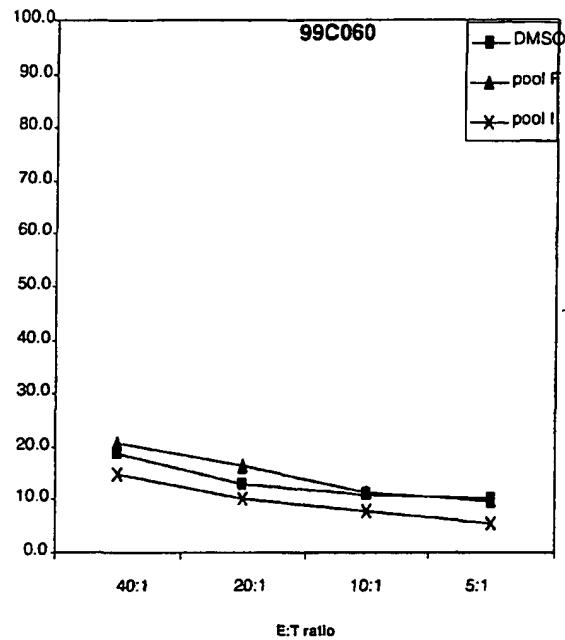
84/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd5-NSmut.

FIG. 18C

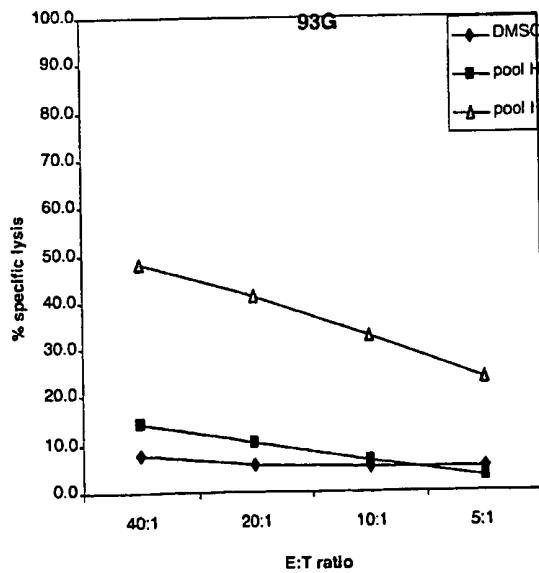
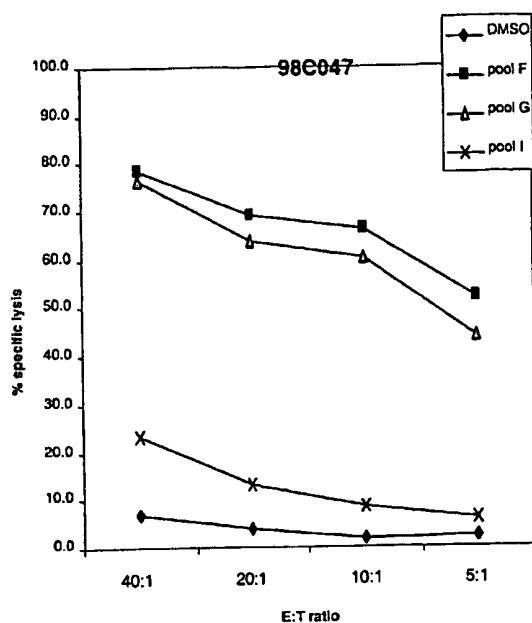
85/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAdS-NSmut

FIG. 18D

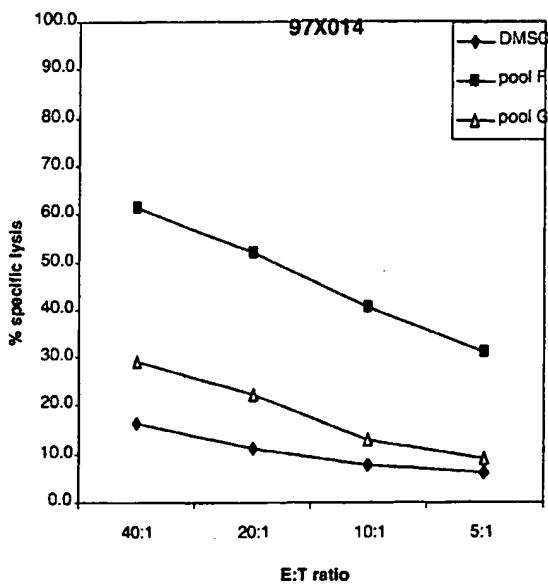
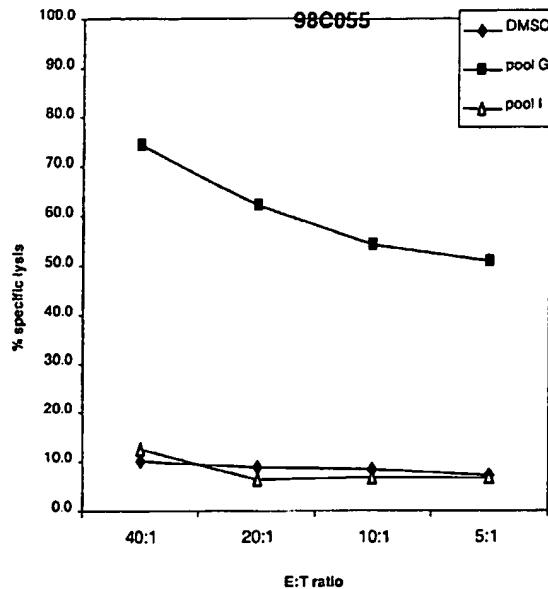
86/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd6-NSmut.

FIG. 18E

87/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd6-NSmut.

FIG. 18F

88/92

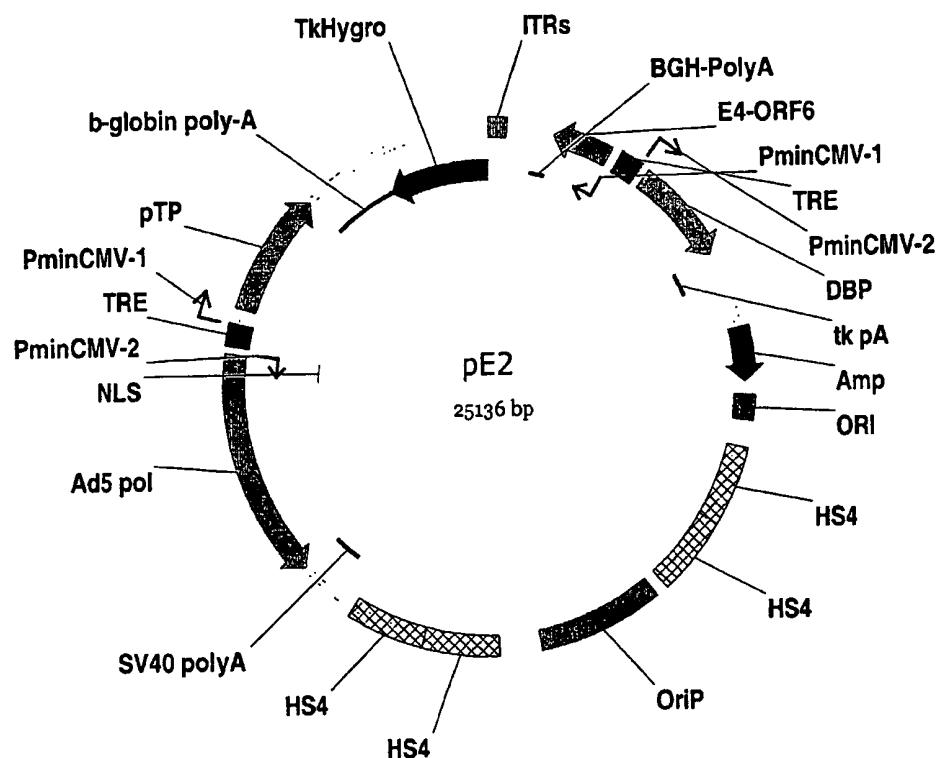


FIG. 19

89/92

1 GCCACCATGG CCCCATCAC CGCCTACAGC CAGCAGACCA GGGGCCTGCT
51 GGGCTGCATC ATCACCAAGCC TGACCGGACG CGACAAGAAC CAGGTGGAGG
101 GAGAGGTGCA GGTGGTGAGC ACCGCTACCC AGAGCTTCTT GGCCACCTGC
151 GTGAACGGCG TGTGCTGGAC CGTGTACAC GGAGCCGAA GCAAGACCC
201 GGCCGGACCC AAGGGCCCTA TCACCCAGAT GTACACCAAT GTGGATCAGG
251 ATCTGGTGGG CTGGCAGGCC CCTCCCGGAG CCAGGAGCCT GACACCTGT
301 ACCTGTGGAA GCAGCGACCT GTACCTGGTG ACACGCCACG CCGATGTGAT
351 CCCCGTGAGG CGCAGGGCG ATTCTCGCGG AAGCCTGCTG AGCCCTAGGC
401 CCGTGAGCTA CCTGAAGGGC AGCAGCGGAG GACCCCTGCT GTGTCCTTCT
451 GGCCATGCCG TGGCATTGGG TCGCGCTGCC GTGTGTACCA GGGGCGTGGC
501 CAAAGCCGTG GATTTGTGC CGTGGAAAG CATGGAGACC ACCATGCGCA
551 GCCCTGTGTT CACCGACAAC AGCTCTCCCC CTGCCGTGCC CCAATCATTC
601 CAGGTGGCTC ACCTGCACGC CCCTACCGGA TCTGGCAAGA GCACCAAGGT
651 GCCCGCTGCC TACGCCGCTC AGGGCTACAA GGTGCTGGTG CTGAACCCCA
701 GCGTGGCCGC TACCCCTGGG TTCGGCGCTT ACATGAGCAA GGCCCATGGC
751 ATCGACCCCA ACATCCGCAC AGGCCTGCAC ACCATCACCA CCGGAGCTCC
801 CGTGACCTAC AGCACCTACG GCAAGTTCTT GGCGATGGA GGCTGCAGCG
851 GAGGAGCCTA CGACATCATC ATCTGCACG AGTGCACAG CACCGACAGC
901 ACCACCATCC TGGCATTGG CACCGTGCTG GATCAGGCCG AAACAGCTGG
951 AGCCAGGCTG GTGGTGCTGG CCACAGCTAC CCCTCCTGGC AGCGTGACCG
1001 TGCCCCATCC CAATATCGAG GAGGTGGCCC TGAGCAACAC AGGCGAGATC
1051 CCCTTCTACG GCAAGGCCAT CCCCATCGAG GCCATCCCGG GAGGCAGGCA
1101 CCTGATCTTC TGCCACAGCA AGAAGAAGTG CGACGAGCTG GCTGCCAAGC
1151 TGAGCGGACT GGGCATCAAC CCCGTGGCCT ACTACAGGGG CCTGGACGTG
1201 TCAGTGATCC CCACCATCGG CGATGTGGTG GTGGTGGCCA CCGACGCCCT
1251 GATGACAGGC TACACCGGAG ACTTCGACAG CGTGATCGAC TGCAACACCT
1301 GCGTGACCCA GACCGTGGAC TTCAGCCTGG ACCCCACCTT CACCATCGAA
1351 ACCACCAACG TGCCTCAGGA TGCTGTGAGC AGGAGCCAGA GGCAGGGACG
1401 CACCGGAAGG GGCAGGGCG GAATTATCG CTTGTGACC CCTGGCGAAA
1451 GGCCCTCTGG CATGTTCGAC AGCAGCGTGC TGTGCGAGTG CTACGACGCT
1501 GGCTGGCTT GGTACGAGCT GACACCCGCT GAAACCAGCG TGCGCCTGCC
1551 CGCTTATCTG AATACCCCTG GCCTGCCGT GTGTCAGGAC CACCTGGAGT

FIG. 20A

90/92

1601 TCTGGGAGAG CGTGTTACA GGACTGACCC ACATCGACGC CCATTCCTG
1651 AGCCAGACCA AGCAGGCTGG CGACAACCTTC CCCTATCTGG TGGCCTATCA
1701 GGCCACCGTG TGTGCTAGGG CCCAAGCTCC ACCTCCTTCA TGGGACCAGA
1751 TGTGGAAGTG CCTGATCCGC CTGAAGCCCA CCCTGCACGG CCCTACCCCT
1801 CTGCTGTACC GCCTGGGAGC CGTGCAGAAC GAGGTGACCC TGACCCACCC
1851 CATCACCAAG TACATCATGG CCTGCATGAG CGCTGATCTG GAAGTGGTGA
1901 CCAGCACCTG GGTGCTGGTG GGAGGCGTGC TGGCCGCTCT GGCTGCCTAC
1951 TGCCTGACCA CCGGAAGCGT GGTGATCGTG GGACGCATCA TCCTGAGCGG
2001 AAGGCCCGCT ATCGTGCCTG ATCGCGAGTT CCTGTACCAAG GAGTCGACG
2051 AGATGGAGGA GTGTGCCAGC CACCTGCCCT ACATCGAGCA GGGCATGCAG
2101 CTGGCCGAAC AGTTCAAGCA GAAGGCCCTG GGCCTGCTGC AGACAGCCAC
2151 CAAACAGGCC GAAGCTGCCG CTCCCGTGGT GGAAAGCAAG TGGAGGGCCC
2201 TGGAGACCTT CTGGGCTAAG CACATGTGGA ACTTCATCTC TGGCATCCAG
2251 TACCTGGCCG GACTGAGCAC CCTGCCTGGC AACCCCGCTA TCGCCAGCCT
2301 GATGGCCTTC ACCGCTAGCA TCACCTCTCC CCTGACCACC CAGAGCACCC
2351 TGCTGTTCAA CATTCTGGC GGATGGGTGG CCGCTCAGCT GGCCCCCTCCT
2401 TCAGCTGCTT CTGCCTTGT GGGCGCTGGC ATTGCCGGAG CCCCTGTGGG
2451 CAGCATTGGC CTGGGCAAAG TGCTGGTGG AATTCTGGCT GGCTATGGCG
2501 CTGGCGTGGC CGGAGCCCTG GTGGCCTTC AAGGTGATGAG CGGAGAGATG
2551 CCCAGCACCG AGGACCTGGT GAACCTGCTG CCTGCCATTC TGAGCCCTGG
2601 AGCCCTGGTG GTGGCGTGG TGTGTGCTGC CATTCTGAGG CGCCATGTGG
2651 GACCCGGAGA GGGCGCTGTG CAGTGGATGA ACCGCCTGAT CGCCTTCGCC
2701 TCTCGCGGAA ACCACGTGAG CCCTACCCAC TACGTGCCTG AGAGCGACGC
2751 CGCTGCCAGG GTGACCCAGA TCCGTGAGCAG CCTGACCATC ACCCAGCTGC
2801 TGAAGCGCCT GCACCAGTGG ATCAACGAGG ACTGCAGCAC ACCCTGCAGC
2851 GGAAGCTGGC TGAGGGACGT GTGGGACTGG ATCTGCACCG TGCTGACCGA
2901 CTTCAAGACC TGGCTGCAGA GCAAGCTGCT GCCCCAATG CCTGGCGTGC
2951 CCTTCTTCTC ATGCCAGCGC GGATACAAGG GCGTGTGGAG GGGCGATGGC
3001 ATCATGCAGA CCACCTGTCC CTGCGGAGCC CAGATCACAG GCCACGTGAA
3051 GAACGGCAGC ATGCGCATCG TGGGCCCTAA GACCTGCAGC AACACCTGGC
3101 ACGGCACCTT CCCCCATCAAC GCCTACACCA CCGGACCCCTG CACACCCAGC
3151 CCTGCTCCCA ACTACAGCAG GGCCCTGTGG AGGGTGGCTG CCGAGGAGTA

FIG. 20B

91/92

3201 CGTGGAGGTG ACCAGGGTGG GAGACTTCCA CTACGTGACC GGAATGACCA
3251 CCGACAACGT GAAGTGTCCC TGTCAGGTGC CCGCTCCCGA ATTTTTTACC
3301 GAAGTGGATG GCGTGCCCT GCATCGCTAT GCCCCCTGCCT GTAGGCCCT
3351 GCTGCGCGAA GAAGTGCACCT TCCAGGTGGG CCTGAACCAG TACCTGGTGG
3401 GCAGCCAGCT GCCCTGCCGAG CCTGAGCCCG ATGTGGCCGT GCTGACCAGC
3451 ATGCTGACCG ACCCCAGCCA CATCACAGCC GAAACCGCTA AAAGGCGCCT
3501 GGCCAGGGGC TCTCCTCCAA GCCTGGCCTC AAGCAGCGCT AGCCAGCTGT
3551 CTGCTCCCAG CCTGAAGGCC ACCTGCACCA CCCACCACGT GAGCCCCGAC
3601 GCCGACCTGA TCGAGGCCAA CCTGCTGTGG CGCCAGGAGA TGGGCGGCAA
3651 CATCACCCGC GTGGAGAGCG AGAACAAAGGT GGTGGTGCTG GACAGCTTCG
3701 ACCCCCTGCG CGCCGAGGAG GACGAGCGCG AGGTGAGCGT GCCCCCGAG
3751 ATCCTGCGCA AGAGCAAGAA GTTCCCCGCT GCCATGCCA TCTGGGCTAG
3801 ACCTGATTAC AACCCCTCCCC TGCTGGAGAG CTGGAAGGAC CCTGATTACG
3851 TGCCTCCAGT GGTGCATGGC TGTCCCTCTGC CTCCCATTAA AGCCCCCTCCT
3901 ATTCCACCTC CTAGGCGCAA AAGGACCGTG GTGCTGACAG AAAGCAGCGT
3951 GAGCTCTGCT CTGGCCGAAC TGGCCACCAA GACCTTTGGC AGCAGCGAGA
4001 GCTCTGCCGT GGACAGCGGA ACAGCCACCG CTCTGCCCTGA CCAGGCCAGC
4051 GACGACGGCG ATAAGGGCAG CGATGTGGAG AGCTATAGCA GCATGCCCTCC
4101 CCTGGAAGGC GAACCTGGCG ATCCCGATCT GAGCGATGGC AGCTGGAGCA
4151 CCGTGAGCGA AGAGGCCAGC GAGGACGTGG TGTGTTGCAG CATGAGCTAC
4201 ACCTGGACAG GCGCTCTGAT CACACCCCTGC GCTGCCGAGG AGAGCAAGCT
4251 GCCCATCAAC GCCCTGAGCA ACAGCCTGCT GAGGCACCCAC AACATGGTGT
4301 ACGCCACCAAC CAGCAGGTCT GCCGGACTGA GGCAGAAAGAA GGTGACCTTC
4351 GACCGCCTGC AGGTGCTGGA CGACCACTAC CGCGATGTGC TGAAGGAGAT
4401 GAAGGCCAAG GCCAGCACCG TGAAGGCCAA GCTGCTGAGC GTGGAGGAGG
4451 CCTGCAAGCT GACCCCCCCC CACAGCGCCA AGAGCAAGTT CGGCTACGGC
4501 GCCAAGGACG TGCGAACCT GAGCAGCAAG GCCGTGAACC ACATCCACAG
4551 CGTGTGGAAG GACCTGCTGG AGGACACCGT GACCCCCATC GACACCACCA
4601 TCATGGCCAA GAACGAGGTG TTCTGCGTGC AGCCCGAGAA GGGCGGCCCGC
4651 AAGCCCGCTC GCCTGATCGT GTTCCCCGAT CTGGGCGTGC GCGTGTGCGA
4701 GAAGATGGCC CTGTACGACG TGGTGAGCAC CCTGCCCTAG GTGGTGATGG
4751 GCTCAAGCTA CGGCTTCCAG TACAGCCCTG GCCAGCGCGT GGAGTTCCCTG

FIG. 20C

92/92

4801 GTGAACACCT GGAAGAGCAA GAAGAACCCC ATGGGTTCA GCTACGACAC
4851 ACGCTGCTTC GACAGCACCG TGACCGAGAA CGACATCCGC GTGGAGGAGA
4901 GCATCTACCA GTGCTGCGAC CTGGCCCTG AGGCCAGGCA GGCCATCAAG
4951 AGCCTGACCG AGCCCTGTA CATCGGAGGC CCTCTGACCA ACAGCAAGGG
5001 ACAGAACTGC GGATAACAGGC GCTGTAGGGC CTCTGGCGTG CTGACCACCA
5051 GCTGTGGCAA CACCCGTGACC TGCTACCTGA AGGCCAGCGC TGCCTGTCGC
5101 GCTGCCAAGC TGCAGGACTG CACCATGCTG GTGAACGCCG CTGGCCTGGT
5151 GGTGATTGT GAAAGCGCTG GCACCCAGGA AGATGCTGCC AGCCTGCGCG
5201 TGTTCACCGA GGCCATGACC AGGTACTCTG CCCCTCCCG AGACCCCCCT
5251 CAGCCGAAT ACGACCTGGA GCTGATCACC AGCTGCTCAA GCAACGTGAG
5301 CGTGGCTCAC GACGCCAGCG GAAAGCGCGT GTACTACCTG ACACCGGATC
5351 CCACCAACCC TCTGGCTCGC GCTGCCTGGG AAACCGCTCG CCATACACCC
5401 GTGAACAGCT GGCTGGCAA CATCATCATG TACGCCCTA CCCTGTGGC
5451 TCGCATGATC CTGATGACCC ACTTCTTCAG CATCCTGCTG GCTCAGGAGC
5501 AGCTGGAGAA GGCCCTGGAC TGCCAGATT ACGGCGCTTG CTACAGCATC
5551 GAGCCCTGG ACCTGCCCA AATCATCGAG CGCCTGCACG GCCTGTCTGC
5601 CTTCAGCCTG CACAGCTACA GCCCTGGCGA AATTAATCGC GTGGCCAGCT
5651 GTCTGCGCAA ACTGGGCGTG CCTCCTCTGC GCGTGTGGAG GCATAGGGCT
5701 AGGAGCGTGA GGGCTAGGCT GCTGAGCCAG GGAGGCAGGG CCGCTACCTG
5751 TGGAAAGTAC CTGTTCAACT GGGCCGTGAA GACCAAGCTG AAGCTGACCC
5801 CTATCCCTGC CGCTAGCCAG CTGGACCTGA GCGGATGGTT CGTGGCTGGC
5851 TACAGCGGAG GCGACATCTA CCACAGCCTG TCTCGCGCTC GCCCTCGCTG
5901 GTTCATGCTG TGCCTGCTGC TGCTGAGCGT GGGCGTGGC ATCTACCTGC
5951 TGCCCAACCG CTAAA

FIG. 20D

**IN THE PCT RECEIVING OFFICE
OF THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant(s):	Merck & Co., Inc		
PCT Serial No.:	To Be Assigned	Case No.:	PCT ITR0015Y
Filing date:	On Even Date Herewith		
For:	HEPATITIS C VIRUS VACCINE		
		Authorized Officer:	US/RO
		To Be Assigned	

Assistant Commissioner of Patents
BOX PCT
Washington, D.C. 20231

**NUCLEOTIDE AND/OR AMINO ACID
SEQUENCE DISCLOSURE, PCT RULE 5.2**

Sir:

As required under PCT Rule 5.2, Applicant respectfully encloses a paper (64 pages) and a computer readable form of the Sequence Listing for the above-identified PCT International Application, filed on even date herewith.

I hereby state that the content of the paper and computer readable forms of the Sequence Listing, submitted in accordance with WIPO and Standard ST.23 and under PCT Rule 13ter.1, respectively, are the same.

Respectfully submitted,

By Sheldon O. Heber
Sheldon O. Heber
Reg. No. 38,179
Attorney for Applicants

Merck & Co., Inc.
P.O. Box 2000
Rahway, NJ 07065-0907
(732) 594-1958

SEQUENCE LISTING

<110> Merck & Co. Inc., and Istituto Di Ricerche Di Biologia Molecolare P.
Angeletti S.P.A.

<120> HEPATITIS C VIRUS VACCINE

<130> ITR0015Y

<150> 60/363,774

<151> 2002-03-13

<150> 60/328,655

<151> 2001-10-11

<160> 17

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1985

<212> PRT

<213> Artificial Sequence

<220>

<223> Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide

<400> 1

Met Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly
1 5 10 15
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
20 25 30
Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys
35 40 45
Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr
50 55 60
Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp
65 70 75 80
Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr
85 90 95
Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
100 105 110
Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
115 120 125
Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
130 135 140
Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys
145 150 155 160
Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met
165 170 175
Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro
180 185 190
Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly
195 200 205

Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr
 210 215 220
 Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly
 225 230 235 240
 Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly
 245 250 255
 Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly
 260 265 270
 Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile
 275 280 285
 Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile
 290 295 300
 Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val
 305 310 315 320
 Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn
 325 330 335
 Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly
 340 345 350
 Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe
 355 360 365
 Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly
 370 375 380
 Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val
 385 390 395 400
 Ile Pro Thr Ile Gly Asp Val Val Val Ala Thr Asp Ala Leu Met
 405 410 415
 Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys
 420 425 430
 Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu
 435 440 445
 Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly
 450 455 460
 Arg Thr Gly Arg Gly Arg Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly
 465 470 475 480
 Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr
 485 490 495
 Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val
 500 505 510
 Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp
 515 520 525
 His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp
 530 535 540
 Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr
 545 550 555 560
 Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro
 565 570 575
 Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr
 580 585 590
 Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn
 595 600 605
 Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met
 610 615 620
 Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly
 625 630 635 640

Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val
 645 650 655
 Ile Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp
 660 665 670
 Arg Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser
 675 680 685
 His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys
 690 695 700
 Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala
 705 710 715 720
 Ala Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp
 725 730 735
 Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly
 740 745 750
 Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe
 755 760 765
 Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe
 770 775 780
 Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala
 785 790 795 800
 Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser
 805 810 815
 Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala
 820 825 830
 Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met
 835 840 845
 Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro
 850 855 860
 Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His
 865 870 875 880
 Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala
 885 890 895
 Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu
 900 905 910
 Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr Ile
 915 920 925
 Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser
 930 935 940
 Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys
 945 950 955 960
 Thr Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro
 965 970 975
 Gln Leu Pro Gly Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys Gly
 980 985 990
 Val Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly Ala
 995 1000 1005
 Gln Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro
 1010 1015 1020
 Lys Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr
 1025 1030 1035 1040
 Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala
 1045 1050 1055
 Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly
 1060 1065 1070

Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro
 1075 1080 1085
 Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg
 1090 1095 1100
 Leu His Arg Tyr Ala Pro Ala Cys Arg Pro Leu Leu Arg Glu Glu Val
 1105 1110 1115 1120
 Thr Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro
 1125 1130 1135
 Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp
 1140 1145 1150
 Pro Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly
 1155 1160 1165
 Ser Pro Pro Ser Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro
 1170 1175 1180
 Ser Leu Lys Ala Thr Cys Thr Thr His His Val Ser Pro Asp Ala Asp
 1185 1190 1195 1200
 Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile
 1205 1210 1215
 Thr Arg Val Glu Ser Glu Asn Lys Val Val Val Leu Asp Ser Phe Asp
 1220 1225 1230
 Pro Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu
 1235 1240 1245
 Ile Leu Arg Lys Ser Lys Lys Phe Pro Ala Ala Met Pro Ile Trp Ala
 1250 1255 1260
 Arg Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp
 1265 1270 1275 1280
 Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala
 1285 1290 1295
 Pro Pro Ile Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu
 1300 1305 1310
 Ser Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly
 1315 1320 1325
 Ser Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro
 1330 1335 1340
 Asp Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr
 1345 1350 1355 1360
 Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser
 1365 1370 1375
 Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val
 1380 1385 1390
 Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys
 1395 1400 1405
 Ala Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu
 1410 1415 1420
 Leu Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly
 1425 1430 1435 1440
 Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp
 1445 1450 1455
 His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val
 1460 1465 1470
 Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro
 1475 1480 1485
 His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn
 1490 1495 1500

Leu Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu
 1505 1510 1515 1520
 Leu Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn
 1525 1530 1535
 Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg
 1540 1545 1550
 Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala
 1555 1560 1565
 Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser
 1570 1575 1580
 Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn
 1585 1590 1595 1600
 Thr Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg
 1605 1610 1615
 Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser
 1620 1625 1630
 Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys
 1635 1640 1645
 Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys
 1650 1655 1660
 Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr
 1665 1670 1675 1680
 Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala
 1685 1690 1695
 Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Ala Ala
 1700 1705 1710
 Gly Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala
 1715 1720 1725
 Ser Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro
 1730 1735 1740
 Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys
 1745 1750 1755 1760
 Ser Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr
 1765 1770 1775
 Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu
 1780 1785 1790
 Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met
 1795 1800 1805
 Tyr Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe
 1810 1815 1820
 Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln
 1825 1830 1835 1840
 Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile
 1845 1850 1855
 Ile Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser
 1860 1865 1870
 Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val
 1875 1880 1885
 Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg
 1890 1895 1900
 Leu Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe
 1905 1910 1915 1920
 Asn Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala
 1925 1930 1935

Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly
 1940 1945 1950
 Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu
 1955 1960 1965
 Cys Leu Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn
 1970 1975 1980
 Arg
 1985

<210> 2
 <211> 5965
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Non-optimized cDNA sequence encoding SEQ. ID. NO.
 1

<400> 2

gccaccatgg	cgcccatcac	ggcctactcc	caacagacgc	ggggcctact	tggttgcattc	60
atcaactagcc	ttacaggccg	ggacaagaac	caggtcgagg	gagaggttca	ggtggtttcc	120
accgcaacac	aatccttct	ggcgcacctgc	gtcaacggcg	tgtgttggac	cgttaccat	180
ggtgctggct	caaagacctt	agccggccca	aaggggccaa	tcacccagat	gtacactaat	240
gtggaccagg	acctcgctgg	ctggcaggcg	ccccccgggg	cgcgttccct	gacaccatgc	300
acctgtggca	gctcagacct	ttacttggtc	acgagacatg	ctgacgtcat	tccggtgcgc	360
ccggggggcg	acagtagggg	gaggctgctc	tccccaggc	ctgtctcta	tttgaaggc	420
tcttcgggt	gtccactgct	ctgcccattcg	gggcacgctg	tgggcatttt	ccgggctgcc	480
gtatgcaccc	gggggggtgc	gaaggcggtg	gactttgtgc	ccgttagatgc	catggaaact	540
actatgcgg	ctccggctt	cacggacaac	tcatcccccc	cggccgtacc	gcagtcat	600
caagtggccc	acctacacgc	tcccactggc	agcggcaaga	gtactaaagt	gccggctgca	660
tatgcagccc	aagggtacaa	gtgtcgctc	ctcaatccgt	ccgttgccgc	taccttaggg	720
tttggggcgt	atatgtctaa	ggcacacgg	attgacccca	acatcagaac	tgggtaagg	780
accattacca	caggcgcccc	cgtcacatac	tctacctatg	gcaagtttct	tgccgatggt	840
ggttgctctg	ggggcgctta	tgacatcata	atatgtatg	agtgcattc	aactgactcg	900
actacaatct	tggcattcg	cacagtctg	gaccaagcgg	agacggctgg	agcgcgctt	960
gtcgtgctcg	ccaccgctac	gcctccggga	tgggtcaccg	tgccacaccc	aaacatcgag	1020
gagggtggccc	tgtctaatac	tggagagatc	cccttctatg	gcaaagccat	ccccattgaa	1080
gccatcagg	ggggaaaggca	tctcatttc	tgtcattcca	agaagaagtg	cgacgagctc	1140
gcccggaaac	tgtcaggcct	cggaatcaac	gctgtggcgt	attaccgggg	gctcgatgt	1200
tccgtcatac	caactatcg	agacgtcgtt	gtcgtggcaa	cagacgctct	gatgacggc	1260
tatacggcg	actttgactc	agtgtatcgac	tgtaaacat	gtgtcaccct	gacagtgcac	1320
ttcagctgg	atcccacctt	caccatttag	acgacgaccg	tgcctcaaga	cgcagtgtcg	1380
cgctcgacg	ggcggggtag	gactggcagg	gttaggagag	gcatctacag	gtttgtact	1440
ccggggagaac	ggccctcg	catgttcgat	tcctcggtcc	tgtgtgagtg	ctatgacgcg	1500
ggctgtgctt	ggtacgagct	caccccccgg	gagacctcg	ttaggtgcg	ggcctacctg	1560
aacacaccag	ggttggccgt	ttgcccaggac	cacctggagt	tctgggagag	tgtttcaca	1620
ggcctaccc	acatagatgc	acacttctt	tcccagacca	agcaggcagg	agacaacttc	1680
ccctacctgg	tagcatacca	agccacgg	tgcgcagg	ctcaggcccc	acctccatca	1740
tggatcaaa	tgtgaaagt	tctcatacg	ctgaaacata	cgtgcacgg	gccaacaccc	1800
ttgtgttaca	ggctgggagc	cgtccaaaat	gaggtcacc	tcaacccaccc	cataaccaaa	1860
tacatcatgg	catgcacgtc	ggctgacctg	gaggtcgtca	ctagcacctg	ggtgctgggt	1920
ggcggaggtcc	ttgcacgtct	ggcccgctat	tgcctgacaa	caggcagtgt	ggtcatgt	1980
ggtaggatta	tcttgccgg	gaggccggct	attgtcccg	acagggagtt	tctctaccag	2040
gagttcgtat	aaatggaaga	gtgcgcctcg	cacccctt	acatcgagca	ggaaatgcag	2100
ctcgccgagc	aattcaagca	gaaagcgctc	gggttactgc	aaacagccac	caaacaagcg	2160

gaggctgctg	ctcccggtgt	ggagtccaaag	tggcgagccc	ttgagacatt	ctgggcgaag	2220
cacatgtgga	atttcatcag	cgggatacag	tacttagcag	gcttatccac	tctgcctggg	2280
aaccccgcaa	tagcatcatt	gatggcattc	acagcctcta	tcaccagccc	gctcaccacc	2340
caaagtaccc	tcctgttaa	catcttgggg	gggtgggtgg	ctgccaact	cgccccccc	2400
agcggcgtt	cggcttcgt	gggcgcggc	atcggccgtg	cggctgttgg	cagcataggc	2460
cttggaaagg	tgcttgttga	cattctggcg	ggttatggag	caggagtggc	cggcgcgtc	2520
gtggccttca	aggcatgag	cgcgagatg	ccctccaccc	aggacctgtt	caatctactt	2580
cctgccatcc	tctctccctgg	cggccctggtc	gtcgggggtcg	tgtgtgcage	aatactgcgt	2640
cgacacgtgg	gtccgggaga	gggggctgtg	cagtggatga	accggctgtat	agcgttgcgc	2700
tcgcgggta	atcatgtttc	ccccacgcac	tatgtgcctg	agagcgcacgc	cgcagcgcgt	2760
gttactcaga	tcctctccag	ccttaccatc	actcagctgc	tgaaaaggct	ccaccagtg	2820
attaatgaag	actgctccac	accgtgttcc	ggctcggtgc	taaggatgt	ttgggactgg	2880
atatgcacgg	tgttactga	cttcaagacc	tggctccatgt	ccaagctctt	gcccgcagcta	2940
ccgggagttc	ctttttctc	gtgccaacgc	gggtacaagg	gagtctggcg	gggagacggc	3000
atcatgcaaa	ccacccgtcc	atgtggagca	cagatcacccg	gacatgtcaa	aaacggttcc	3060
atgaggatcg	tcggccctaa	gacctgcaggc	aacacgtggc	atggaaacatt	ccccatcaac	3120
gcatacacca	cgggccctg	cacacccttc	ccagcgccaa	actattcttag	ggcgctgtgg	3180
cgggtggccg	ctgaggagta	cgtggaggtc	acgggggtgg	gggatttcca	ctacgtgacg	3240
ggcatgacca	ctgacaacgt	aaagtgcaca	tgccaggttc	cggctccctga	atttctcactg	3300
gaggtggacg	gagtgcgtt	gcacaggtac	gctccggcg	gcaggccctt	cctacggag	3360
gaggttacat	tccaggtcgg	gctcaaccaa	tacctgggtt	ggtcacagct	accatgcgag	3420
cccgaaacccg	atgttagcgt	gctcacttcc	atgctcacccg	accctccca	catcacagca	3480
gaaacggcata	agcgttaggtt	ggccaggggg	tctcccccct	ccttggccag	cttttcagct	3540
agccagttgt	ctgcgccttc	cttgaaggcg	acatgcacta	cccacatgt	ctctccggac	3600
gctgacccatca	tcgaggccaa	cctctgtgg	cggcaggaga	tggggggaa	catcacccgc	3660
gtggagtcgg	agaacaagggt	ggtagtcctg	gactcttcg	accgccttcg	agcggaggag	3720
gatgagaggg	aagtatccgt	tccggcgag	atcctgcgga	aatccaagaa	gttcccccga	3780
gcgtatccca	tctggcgcg	cccgattac	aaccctccac	tgttagagtc	ctggaaaggac	3840
ccggactacg	tccctccgg	ggtcacggg	tggccgttgc	cacctatcaa	ggcccccctca	3900
ataccaccc	cacggagaaa	gaggacgggtt	gtcctaacag	agtcctccgt	gtttctgccc	3960
ttagcgagc	tcgtactaa	gacccctcgcc	agctccgaat	catcgccgt	cgacagcggc	4020
acggcgacc	cccttcctga	ccaggcctcc	gacgacggtg	acaaaggatc	cgacgtttag	4080
tcgtactcct	ccatggccccc	ccttgagggg	gaacgggggg	accccgatct	cagtgacggg	4140
tcttggtcta	ccgtgagcga	ggaagctagt	gaggatgtcg	tctgtctgctc	aatgtccctac	4200
acatggacag	gchgcttgc	cacccatgc	gctgcccgg	aaagcaagct	gcccacatcaac	4260
gcgttgcgca	actctttgc	gcccacccat	aacatggttt	atcccaacac	atctcgacgc	4320
gcaggccctgc	ggcagaagaa	ggtcacctt	gacagactgc	aagtccctgg	cgaccactac	4380
cgggacgtgc	tcaaggagat	gaaggcgaag	gctccacag	ttaaggctaa	actcctatcc	4440
gttagaggaag	cctgcaagct	gaccccccca	cattcgccca	aatccaagtt	tggctatggg	4500
gcaaaggacg	tccggaaacct	atccagcaag	gcccgttaacc	acatccactc	cgtgtggaa	4560
gacttgcgg	aagacactgt	gacaccaatt	gacaccacca	tcatggcaaa	aatatggagtt	4620
ttctgtgtcc	aaccagagaa	aggaggccgt	aaggccagccc	gccttacatgt	attcccgat	4680
ctggggatcc	gtgtatgcga	gaagatggcc	ctctatgtat	tgtctccac	ccttcctcag	4740
gtcgtatgg	gctcctcata	cggattccag	tactctcctg	ggcagcgcagt	cgagttccctg	4800
gtgaataacct	ggaatcaaa	aaaaaaacccc	atgggtttt	catatgacac	tgcgtgtttc	4860
gactcaacgg	tcaccgagaa	cgacatccgt	gttggaggat	caatttacca	atgttgcac	4920
ttggcccccg	aagccagaca	ggccataaaaa	tcgctcacag	agcggcttta	tatcggggg	4980
cctctgacta	attccaaaagg	gcagaactgc	ggttatgcgc	ggtgcgcgc	gagcggcgtg	5040
ctgacgacta	gctgcggtaa	caccctcaca	tgttacttgc	aggcctctgc	agcctgtcg	5100
gctgcgaagc	tccaggactg	cacgtatgc	gtgaacggcg	ccggcccttgc	cgtttatctgt	5160
gaaagcgcgg	gaacccaaaga	ggacggggcg	agccctacgc	tttcacggg	ggctatgact	5220
aggtactctg	cccccccccgg	ggacccccc	caaccagaat	acgacttgg	gctgataaca	5280
tcatgttcc	ccaaatgtgtc	ggtcgcaccc	gatgcacatgc	gcaaaagggt	gtactacctc	5340
acccgtgatc	ccaccaccccc	cctcgacagg	gctgcgtggg	aaacagctag	acacactcca	5400
gttaactcct	ggctaggcaa	cattatcatg	tatgcgcaca	ctttgtggc	aaggatgatt	5460

ctgatgactc acttcttctc catccttcta gcacaggagc aacttgaaaa agccctggac	5520
tgccagatct acggggcctg ttactccatt gagccacttg acctaccta gatcattgaa	5580
cgactccatg gccttagcgc atttcaactc catagttact ctccaggtga gatcaatagg	5640
gtggcttcat gcctcaggaa acttgggta ccacccttgc gagtctggag acatgggcc	5700
aggagcgccc gcgctaggct actgtccccag gggggaggg ccggcacttg tggcaagtac	5760
ctcttcaact gggcagtgaa gaccaaaactc aaactctactc caatccggc tgcgtccag	5820
ctggacttgtt ccggctgggt tacagcgggg gagacatata tcacagcctg	5880
tctcgtgccc gaccggctg gttcatgctg tgcctactcc tactttctgt agggtaggc	5940
atctacctgc tcccccaacctg ataaa	5965

<210> 3
<211> 5965
<212> DNA
<213> Artificial Sequence

<220>
<223> Optimized cDNA encoding SEQ ID NO: 1

<400> 3	
gccaccatgg ccccatcac cgcctacagc cagcagaccc gcccctgct gggctgcac	60
atcaccagcc tgacccggcc cgacaagaac cagggtggagg gcgagggtca ggtggtgagc	120
accggccaccc agagtttctt ggccacctgc gtgaacggcg tttgtctggac cgttaccac	180
ggccgcggca gcaagacccct ggccggccccc aaggggccca tcacccagat gtacaccaac	240
gtggaccagg accttggggg ctggcaggcc ccccccggcg cccgcagcct gacccctgc	300
acctgcgcga gcagcgaccc ttatctgggtt accccggcactg ccgcacgtat cccctgtgc	360
cgccgcgcgc acagccgcgg cagcctgtt agccccggcc ccgtgagcta cctgaaggc	420
agcagcgccg gccccctgct gtgcggccgc ggccacgc tgggcattt cccgcgcgc	480
gtgtgcaccc gccgcgtggc caaggccgtg gacttcgtgc ccgtggagag catggagacc	540
accatgcgcgca gccccgtgtt caccgacaaac agcagccccc ccgcgcgtgc ccagagcttc	600
cagggtggccc acctgcacgc ccccaaccggc agcggcaaga gacccaagggt gcccgcgc	660
tacggccccc agggttacaa ggtgtctggg ctgaacccca gctgtggccgc caccctggc	720
tccggccct acatgagcaa ggcccacggc atcgacccca acatccgcac cggcgtgc	780
accatcacca ccggccccc cgtgacctac agcacttacg gcaagttcct ggccgacggc	840
ggctgcagcg gccgcgccta cgacatcatc atctgcgacg atgtccacag caccgacagc	900
accaccatcc tgggcattcggtt caccgtgtt gaccaggccg agaccgcgcg cggccgcctg	960
gtgggtctgg ccaccgcac ccccccggc agcgtggccg tggcccaccc caacatcgag	1020
gagggtggccc tgagcaacac cggcgagatc cccttctacg gcaaggccat ccccatcgag	1080
gccccatccgcg gccggccgc cctgatcttc tgccacagca agaagaagtg cgacgagctg	1140
gcccggcaacg tgagcgccct gggcatcaac gcccgtggctt actaccgcgg cctggacgtg	1200
agcgtatcc ccaccatcggtt cggacgtgggtt gtgggtggcca cggacgcctt gatgacccggc	1260
tacaccggcg acttcgacag cgtgatcgac tgcaacaccc ggtgtggccca gaccgtggac	1320
tccagcttgg accccacccctt caccatcgatg accaccacccg tggcccaggaa cccgcgtgagc	1380
cgcagccacg gccgcggccgc caccggccgc ggcgcgcgc gcatctaccg cttcggtacc	1440
ccggcgagc gcccacggc catgttgcac agcagcgatc tggcgtggatg ctacgacgc	1500
ggctgcgcctt ggtacgagct gaccccccggc gagaccacggc tggcgcgtgc cgcctaccc	1560
aacacccccc gctgtccctt gtggcaggac cacctggatg tggggagag cgtgttacc	1620
ggccgtaccc acatcgacgc ccacttccctt agccagacca agcaggccgg cgacaaatcc	1680
cccttacccgg tggcttacca ggccacccgtt tgccgcgcgc cccaggccccc ccccccggc	1740
tggggaccaga tggtaatgtt cctgatccgc ctgaacccca ccctgcacgg cccacccccc	1800
ctgctgtacc gcctggccgc cgtgcagaac gaggtgaccc tgaccaccc cataccaag	1860
tacatcatgg cctgcgttgc cggccacccg gagggtggta ccacgcacccg ggtgttgggt	1920
ggccggcgatc tggccgcctt ggccgcctac tggcgttgcacca cccgcgcgtt ggtgttacc	1980
ggccgcatac tcctgagccgg cccggccccc atcgtcccc accgcgcgtt cctgttacc	2040
gagttcgacg agatggagga tggcgcgcacg cacccgcctt acatcgacca gggcatgcag	2100
ctggccgagc agttcaagca gaaggccctt ggcctgtgc agaccgcac caagcaggcc	2160

gaggccgccc cccccgtggt ggagagcaag tggcgccccc tggagacctt ctggggccaag 2220
 cacatgtgga acttcatcag cggcatccag tacctggccg gcctgagcac cctgcccggc 2280
 aaccccgcca tcgccagcct gatggccctt accgcacagca tcaccagccc cctgaccacc 2340
 cagagcaccc tgctgttcaa catcctgggc ggctgggtgg ccgcccagct ggccccccccc 2400
 agcgccgcca gcgccttcgt gggcgccggc atcgccggcg ccgcccgtggg cagcatggc 2460
 ctgggcaagg tgctgggtggc catcctggcc ggctacggcg cggcgctggc cggcgccctg 2520
 gtggcccttca aggtgatgag cggcgagatg cccagcaccc aggacctggt gaacctgctg 2580
 cccgccatcc tgagccccgg cggccctgggt gtgggctggg tttgcgcgc catcctgcgc 2640
 cgccacgtgg gccccggcga gggcgccgtg cagtggatgat accgcctgtat cgccttcgcc 2700
 agccgcggca accacgtgag ccccacccac tacgtgtcccc agagcgacgc cggcgccccc 2760
 gtgacccaga tcctgagcag cctgaccatt acccagctgc tgaagcgccct gcaccagtg 2820
 atcaacgagg actgcagcac cccctgcagg ggcagctggc tgcgcgcacgt gtgggactgg 2880
 atctgcaccc tgctgaccga cttcaagacc tggctgcaga gcaagctgt gccccagctg 2940
 cccggcggtgc ccttettccat ctggccagcgc ggctacaagg gcgtgtggcg cggcgacggc 3000
 atcatgcaga ccacccgttccat ctcggccgc cagatcaccg gccacgtgaa gaacggcagc 3060
 atgcgcaccc tgggcccaa gacctgcagg aacacccgttccat acggcacctt ccccatcaac 3120
 gcctacacca cccggccctg caccggcaggc cccggcccccactacagccg cggccctgtgg 3180
 cgcgtggcc ccgaggagta cgtggagggtt acccgcgtgg gcgacttcca ctacgtgacc 3240
 ggcatgcacca cgcacaacgt gaagtggccc tgccagggtgc cggccccca gtttttacc 3300
 gaggtggacg gcgtgcgcct gcaccgttac gccccccctt gctgcgcag 3360
 gaggtgaccc tccagggtggg cctgaaccag tacctgggtgg gcagccagct gcccctgcg 3420
 cccgagcccg acgtggccgt gctgaccaggc atgctgaccg accccagccatcaccggcc 3480
 gagaccgcacca agcggccctt gggccggc agccccccca gcctggccag cagcagcgcc 3540
 agccagctga gcccggccag cctgaaggcc acctgcacca cccaccacgt gagccccgac 3600
 gcccacctga tcgaggccaa cctgctgtgg cggcaggaga tggcgccaa catcaccggc 3660
 gtggagagcg agaacaagggt ggtggtgctg gacagcttcg accccctgcg cggcgaggag 3720
 gacgagcgcc aggtgagcgt gggccggag atcctgcgc acagcaagaa gttttttcc 3780
 gccatgcacca tctggcccg ccccgactac aaaaaaaaaa tttttttttt gttttttttt 3840
 cccgactacg tgccccccgt ggtgcacggc tgccccctgc ccccccataa gggccccccc 3900
 atccccccccc cccggccacca ggcgcaccgtg gtgtgaccg agacgacgtt gggccggcc 3960
 ctggccgagc tggccacccaa gacccctgcg acgacggaga gcaacggccgt ggacacggc 4020
 accggccaccc ccctggccgaa ccaggccaggc gacgacggcg acaacggccag cgacgtggag 4080
 agctacacca gcatggccccc cctggaggcc gggccggcc accccgaccc gggccggcc 4140
 agctggagca cctgtggccggc ggaggccaggc gggacgtgg ttgtgctgcag catgagctac 4200
 acctggaccg ggcgcctgtat caccggccgc gggccggagg agacgacgtt gggccggcc 4260
 gcccctgagca acaggctgtt ggcgcacccaa aacatgggtt acggccacccac cggccgcac 4320
 gcccggccgc gcccggccggaa ggtgacccctt gaccgcctgc aggtgctggaa cggccactac 4380
 cggcgcacgtgc tgaaggagat gaaggccaa ggcgcacccg tgaaggccaa gtttgcgttgc 4440
 gtggaggagg cctgcaagct gaccccccacacggccca agacgacgtt cggctacggc 4500
 gccaaggacg tgcgcacccctt gggccggcc gggccggcc acatccaccc cgtgtggaa 4560
 gacctgctgg aggacaccgt gaccccccattt gacaccacca tcatggccaa gacgagggtg 4620
 ttctgcgtgc agcccgagaa gggccggccgc aaccccccggc gctgtatcgt gttttttcc 4680
 ctggccgtgc gctgtgtgcgaa gaagatggcc ctgtacgcgac tttttttttt gggccggcc 4740
 gtgggtatgg gcacggacta cggccctccat tacaggcccg gggccggccgt gggatccctg 4800
 gtgaacaccc ggaagagccaa gaagaacccccccatgggttca gtttgcgttgc cccctgttgc 4860
 gacacccggc tgaccggagaa cggccatccgc gtttgcgttgc gtttgcgttgc gtttgcgttgc 4920
 ctggcccccgg aggccggccca gggccatccgc acggccggcc acggccggcc gggccggcc 4980
 cccctgacca acacccggccca cggccatccgc gggccggcc gggccggcc gggccggcc 5040
 ctgaccacca gtttgcgttgc cccctgttgc gtttgcgttgc gtttgcgttgc gtttgcgttgc 5100
 gcccggccggaa tgcggccggcc cccctgttgc gtttgcgttgc gtttgcgttgc gtttgcgttgc 5160
 gagagccggc gcacccggccca gggccatccgc acggccggcc acggccggcc gggccggcc 5220
 cgctacaccc ccccccggccca cggccatccgc gggccggcc gggccggcc gggccggcc 5280
 agctgcaccc gcaacgttgg gtttgcgttgc gggccatccgc gggccggcc gggccggcc 5340
 accccggccacc ccacccggccca cccctgttgc gggccatccgc gggccggcc gggccggcc 5400
 gtgaacaccc gggccatccgc gggccatccgc gggccatccgc gggccatccgc gggccatccgc 5460

ctgatgaccc acttcttcag catcctgctg gcccaggagc agctggagaa ggccctggac	5520
tgccagatct acggcgctg ctacagcatc gagccctgg acctgccccca gatcatcgag	5580
cgccctgcacg gcctgagcgc cttcagcctg cacagctaca gccccggcga gatcaaccgc	5640
gtggccagct gcctgcgcaa gctgggcgtg cccccctgc gcgtgtggcg ccaccgcgccc	5700
cgcagcgtgc ggcggccct gctgagccag ggcggccgag ccccacctg cggcaagtac	5760
ctgttcaact gggccgtgaa gaccaagctg aagctgaccc ccatccccgc cgccagccag	5820
ctggacctga gcggctggtt cgtggccggc tacagcggcg gcgacatcta ccacagcctg	5880
agccgcgccc gcccccgctg gttcatgctg tgcctgctgc tgctgagcgt gggcgtggc	5940
atctacctgc tgcccaaccg ctaaa	5965

<210> 4
<211> 37090
<212> DNA
<213> Artificial Sequence

<220>
<223> MRKAd6-NSmut nucleic acid

<400> 4	
catcatcaat aataaacctt attttggatt gaagccaata tgataatgag ggggtggagt	60
ttgtgacgtg. ggcggggcg tgggaacggg gcgggtgacg tagtagtgtg gcgaaagtgt	120
gatgttgc当地 gtgtggcgga acacatgtaa gcgacggatg tggcaaaaagt gacgttttg	180
gtgtgc当地 gtgtacacag gaagtgacaa ttttcgc当地 gttttaggc当地 gatgtttag	240
taaaatggg cgtaaccgag taagatgg ccatttc当地 gggaaaaactg aataagagga	300
agtgaatct gaataatttt gtgttactca tagcgc当地 tatttgctaa gggccgc当地	360
gactttgacc gtttacgtgg agactcgccc aggtttttt ctcagggtt ttccgc当地	420
ccggtaaag ttggc当地 attattatag gcggccgc当地 tccattgc当地 acgttgc当地	480
catatcataa tatgtacatt tatattggct catgtccaaac attaccgc当地 tggc当地	540
gattattgac tagtattaa tagaatcaa ttacgggctc attagttcat agcccatata	600
tggagttccg cgttacataa ctacggtaa atggccgc当地 tggctgaccg cccaaacgacc	660
cccgccatt gacgtcaata atgacgtatg ttcccatagta aacgccaata gggacttcc	720
attgacgtca atgggtggag tatttacggt aaactgc当地 ctggc当地 catcaagtgt	780
atcatatgccc aagtacgc当地 cctattgacg tcaatgacgg taaatggccc gcctggc当地	840
atgccc当地 catgaccta tgggacttcc ctactggca gtacatctac gtattagtca	900
tcgctattac catgtgatg cgggacttcc agtacatcaa tggc当地 tggctgaccg	960
actcacgggg atttcaagt ctccacccca ttgacgtca tgggacttcc tttggc当地	1020
aaaatcaacg ggacttccaa aatgtc当地 acaactccgc cccattgacg caaatggc当地	1080
gtaggc当地 acgggtggag gtctatataa gcagagctcg ttttagtgaac cgtc当地	1140
cctggagacg ccattccacgc tggggacttcc tccatagaag acaccgggac cgatccacgc	1200
tccgc当地 ggaacgggtc当地 attggaaacgc ggatttcccg tggcaagagtg gagatctgc当地	1260
accatggc当地 ccattccacgc ctactcccaa cagacgc当地 gctacttcc ttgc当地	1320
actagc当地 caggccggga caagaaccag gtc当地 gggag aggttc当地 ggttccacc	1380
gcaacacaat ctttctggc gacgtgc当地 aacggc当地 gtggaccgt ttaccatgg	1440
gctggctca agacccctac cggcccaaag gggccaaatca cccagatgtca cactaatgtca	1500
gaccaggacc tcgtc当地 gcaaggc当地 cccggggc当地 gttc当地 accatgc当地	1560
tgtggc当地 cagacccctta ttggctc当地 agacatgtca agtcttcc ggtgc当地	1620
cgggccgaca gtaggggag cctgtctcc cccaggc当地 ttc当地 ttaccatgg	1680
tcggggtggc当地 cactgtctac cc当地 tggggggg caccgtgtgg gcatcttccg ggctgc当地	1740
tgcacccccc ggggtgc当地 ggccggglocal gggccggc当地 tagagtccat ggaaactact	1800
atgc当地 gggttccac ggacaactca tccccccgg cctgaccgc当地 gtc当地	1860
gtggccacc tacacgtcc cactggc当地 ggcaagagta ctaaaatgtcc ggctgc当地	1920
gcagcccaag ggtacaagggt gtc当地 ccttcc aatccgtccg ttggcc当地 acatgtcc	1980
ggggc当地 ttttccgata ttttccgata gaccccaaca tcagaactgg ggtaaggacc	2040
attaccacag ggc当地 cccctgtc cacataactca acctatggca agtttcttcc gcatggggt	2100
tgctctgggg gcgcttcatgca catcataataa tgtgatgagtttcc gcatgtccact	2160

acaatcttgg gcatcgac agtcctggac caagcggaga cggctggagc gcggcttgc 2220
 gtgctcgcca ccgtacgac tccgggatcg gtcaccgtgc cacacccaaa catcgaggag 2280
 gtggccctgt ctaactactgg agagatcccc ttctatggca aagccatccc cattgaagcc 2340
 atcagggggg gaaggcatct cattttctgt cattccaaga agaaatgcga cgagctcgcc 2400
 gcaaagctgt caggcctcg aatcaacgt gtggcgtatt accgggggtt cgatgttcc 2460
 gtcataccaa ctatcgaga cgtcggtgtc gtggcaacag acgtctgtat gacgggttat 2520
 acgggcgact ttgactcagt gatcgactgt aacacatgtc tcacccagac agtcgacttc 2580
 agcttggatc ccacccac cattcgagacg acgaccgtgc ctcagacgc agtgtcgcc 2640
 tcgcagcggc gggtaggac tggcagggtt aggagaggca ttcacaggtt tgtgactccg 2700
 ggagaacggc cctcggtcat gttcgattcc tcggctctgt gtgagtgcta tgacgcggc 2760
 tttgttggt acgagctac ccccgccgag acctcggtt ggttgcggc ctacccgtac 2820
 acaccagggt tgcccggtt ccaggaccac ctggagttt gggagagtgt tttcacaggc 2880
 ctcacccaca tagatgcaca cttcttgcac cagaccaagc aggcaggaga caacctcccc 2940
 tacctggtag cataccaaacg cacgggtgtgc gccagggtctc agggcccccacc tccatcatgg 3000
 gatcaaatgt ggaagtgtct catacggtct aaacccatgc tgacccggcc aacaccctt 3060
 ctgtacaggc tgggagccgt caaaaatgag gtcacccctca cccaccccat aaccaaatac 3120
 atcatggcat gcatgtcgcc tgacctggag gtcgtcacta gcacctgggt gctggtggc 3180
 ggagtccctg cagctctggc cgcgtattgc ctgacaacag gcagtgttgtt cattgtgggt 3240
 aggattatct tgcgggag gccggctatt gttcccgaca gggagtttctt ctaccaggag 3300
 ttcgatgaaa tggaaagatg cgcctcgac cttccattaca tcgagcaggaaatgcagctc 3360
 gcccggcaat tcaagcagaa agcgctcggtt tactgcaaa cagccaccaa acaagcggag 3420
 gctgctgtc ccgtgggttgc gtccaaatggc cgagccctt agacattctg ggcgaagcac 3480
 atgtggaaatt tcatacgacg gatacgtac ttagcaggct tatccactct gcctgggaaac 3540
 cccgcaatag catcattgtat ggcattcaca gcctctatca ccagcccgct caccacccaa 3600
 agtaccctcc tggtaacat cttgggggggg tgggtggctg cccaactcgc cccccccagc 3660
 gcccgttccg cttctggggc cggccggcattc gcccgtgcgg ctgttggcag cataggcctt 3720
 gggaaagggtc ttgtggacat tctgggggtt tatggacgag gagtggccgg cgcgtcg 3780
 gccttcaagg tcatgacggc cgagatgccc tccacccggg acctggtcaa tctacttcc 3840
 gcccattctt ctccctggcgc cttgggtgtc ggggtgtgt gtcagcaat actgcgtcga 3900
 cacgtgggtc cggggagaggg ggctgtgcag tggatgaaacc ggcgtatagc gtccgtctg 3960
 cggggtaatc atgtttcccc caacgactat gtgcctgaga gcgacgcgc agcgcgtt 4020
 actcagatcc tctccagcct taccatcact cagctgctga aaaggctcca ccagtggatt 4080
 aatgaagact gtcacacacc gtgttccggc tctgtggctaa gggatgtttt ggactggata 4140
 tgcacgggtt gactgactt caagacactgg ctccagtcac agtcctgtcc gcagctaccg 4200
 ggagtccctt ttttctgtc ccaacgcggg tacaaggggat tctggccggg agacggcatc 4260
 atgcaaaacca cctggccatg tggagcacaatc accccggac atgtcaaaaaa cggttccatg 4320
 aggatecgctg ggcctaagac ctgcagcaac acgtggcatg gaacattccc catcaacgca 4380
 tacaccacgg gcccgtcacc accctctcca ggcacccaaact attctaggc gctgtggcgg 4440
 gtggccgtc aggagatcgt ggaggtcagc cgggtggggg attccacta cgtgacgggc 4500
 atgaccactg acaacgtaaa gtgcctatgc caggttccgg ctctctgtat cttcacggag 4560
 gtggacggag tgcgggtgc caggtacgt cggcgtgc ggcctctctt acggggaggag 4620
 gttacattcc aggtcgggtt caaccaatac ctgggtgggtt cacagctacc atgcgagccc 4680
 gaaccggatg tagcagtgtc cacttccatg ctcacccggacc cctccacat cacagcagaa 4740
 acggctaagc gtaggttggc cagggggatc ccccccctt tggccagctc ttcaagctagc 4800
 cagttgtctg cgccttcctt gaaggcgaca tgcactaccc accatgtctc tccggacgt 4860
 gacctcatcg aggccaaacct cctgtggcgg caggagatgg gccggaaacat caccggcgt 4920
 gagtcggaga acaagggtt gtcctggac tctttcgacc cgttgcggc ggaggaggat 4980
 gagagggaaat tatccgttcc ggcggagatc ctgcggaaat ccaagaagtt ccccgacgc 5040
 atgcccattt gggcgccccc ggattacaac cctccactgt tagagtccctg gaaggaccgc 5100
 gactacgtcc ctccgggttgc gacgggtgc cgggtgccac ctatcaaggc ccctccaata 5160
 ccacccatcg ggagaaaagag gacgggtgc ctaacagat cctccgtgtc ttctgcctt 5220
 gcgagctcg ctactaagac ttccggcagc tccgaatcat cggccgtgc cagcggcagc 5280
 gcgaccgcggcc ttccgtacca ggcctccgac gacgggtgaca aaggatccga ctttgagtcg 5340
 tactcctcca tgccccccct tgagggggaa ccggggggacc ccatctcag tgacgggtct 5400
 tggtctaccg tgagcggagga agctagtgag gatgtcgat gctgctcaat gtcctacaca 5460

tggacaggcg	ccttgatcac	gccatgcgt	gcggagggaaa	gcaagctgcc	catcaacgcg	5520
ttgagcaact	ctttgctgac	ccaccataac	atggttatg	ccacaacatc	tcgcagcgca	5580
ggcctgcggc	agaagaagg	caccttgcac	agactgcaag	tcctggacga	ccactaccgg	5640
gacgtgctca	aggagatgaa	ggcgaaggcg	tccacagtta	aggctaaact	cctatccgt	5700
gaggaaggct	gcaagctgac	gccccacat	tcggccaaat	ccaagttgg	ctatgggca	5760
aaggacgtcc	ggaacctatac	cagcaaggcc	gttaaccaca	tccactccgt	gttgaaggac	5820
ttgcttggaa	acactgtgac	accaattgac	accaccatca	tggcaaaaaaa	tgaggtttc	5880
tgtgtccaa	cagagaaagg	aggccgtaa	ccagcccgc	ttatcgatt	cccagatctg	5940
ggagtcgt	tatgcgagaa	gatggccctc	tatgtgtgg	tctccaccct	tcctcaggc	6000
gtgtatggct	cctcatacgg	attccagta	tctccctggc	agcgagtcga	gttccttgg	6060
aataccttga	aatcaaagaa	aaaccccatg	ggctttcat	atgacactcg	ctgtttcgac	6120
tcaacggtca	ccgagaacga	catccgttt	gaggagtca	tttaccaatg	ttgtgacttg	6180
gcccccaag	ccagacaggc	cataaaatcg	ctcacagagc	ggctttatat	cgggggtcct	6240
ctgactaatt	caaaaggcg	gaaactgcgt	tatcggcggt	gccgcgcgag	cggcgtgctg	6300
acgactagct	gcggtaacac	cctcacatgt	tacttgaagg	cctctgcagc	ctgtcgagct	6360
gCGAAGCTCC	aggactgcac	gatgtctgt	aacgcggcc	gccttgcgt	tatctgtgaa	6420
agcgcggaa	cccaagagga	cgcggcgagc	ctacgagtct	tcacggaggc	tatgactagg	6480
tactctgtcc	ccccgggaa	cccgcggcaa	ccagaatacg	acttggagct	gataacatca	6540
tgttcccca	atgtgtcggt	cgccccacat	gcatcaggca	aaagggtgt	ctacccatc	6600
cgtgatccca	ccacccccc	cgcacgggt	gcgtggaaa	cagctagaca	cactccagtt	6660
aactctggc	taggaacat	tatcatgtat	gcgcccactt	tgtgggcaag	gatgattctg	6720
atgactca	tcttctccat	ccttctagca	caggagcaac	ttgaaaaaagc	cctggactgc	6780
cagatctac	gggcctgtt	ctccatttag	ccacttgacc	tacctcagat	cattgaacga	6840
ctccatggcc	ttagcgcatt	ttcactccat	agttactctc	caggtgagat	caatagggtg	6900
gcttcatgcc	tcagaaact	tgggttacca	cccttgcag	tctggagaca	tcgggcccagg	6960
agcgtccgc	ctaggctact	gtcccagggg	gggaggggcc	ccacttgcgt	caagtacctc	7020
ttcaactggg	cagtgaagac	caaactcaaa	ctcactccaa	tccggctgc	gtcccagctg	7080
gacttgtccg	gctgggtcg	tgctggttac	agcggggag	acatatata	cagcctgtct	7140
cgtccccgac	cccgctgggt	catgtgtgc	ctactctac	tttctgttag	ggttaggcac	7200
tacctgctcc	ccaacccgta	aatctagagc	tgtgccttct	agttgcac	catctgtgt	7260
ttgccccctcc	cccgctgctt	ccttgaccct	ggaaggtgcc	actcccactg	tcctttctta	7320
ataaaaatgag	gaaattgtat	cgcattgtct	gagtaggtgt	cattctattt	tgggggttgg	7380
ggtggggcag	gacagcaagg	gggaggattt	ggaagacaat	aggaggcatg	ctggggatgc	7440
ggtgggctct	atggccgate	ggcgcgcgt	actgaaatgt	gtgggcgtgg	cttaagggtg	7500
ggaaagaata	tataagggtt	gggtctttag	tagtttgt	tctgttttgc	agcagccgc	7560
gcccgcattga	gcaccaactc	gtttgatgga	agcattgt	gctcatattt	gacaacgcgc	7620
atgccccat	gggcgggggt	gcgtcagaat	gtgatggct	ccagcattga	tgtcgcccc	7680
gtcctgcccc	caaactctac	tacctgacc	tacgagaccg	tgtctggAAC	gccgttggag	7740
actgcagcct	ccgcgcgcgc	ttcagccgt	gcagccac	ccgcgggat	tgtgactgac	7800
tttgcttcc	tgagcccgct	tgcaaggcgt	gcagttccc	gtcataccgc	ccgcgtatgac	7860
aagttgacgg	ctctttggc	acaattggat	tcttgcaccc	ggaacttaa	tgtcgttct	7920
cagcagctgt	tggatctgc	ccagcagggt	tctgcctga	agcttcctc	ccctcccaat	7980
gcggtttaaa	acataaataa	aaaaccagac	tctgtttga	tttggatcaa	gcaagtgtct	8040
tgctgtctt	atttagggtt	tttgcgcgc	cggtaggccc	gggaccagcg	gtctcgtcg	8100
ttgaggggtcc	tgtgtatTTT	ttccaggac	tggtaaagg	gactctggat	gttcagatac	8160
atgggcataa	gcccgtctc	gggggtggagg	tagcaccact	gcagagcttc	atgctgcggg	8220
gtgggttgtt	agatgatcca	gtcgttagcag	gagcgttgg	cgtgggcct	aaaaatgtct	8280
ttcagtagca	agctgattgc	cagggcagg	cccttgggt	aagtgtttac	aaagcgttta	8340
agctggatg	ggtgcatacg	tggggatatg	agatgatct	tgactgtat	tttttaggtt	8400
gctatgttcc	cagccatatac	cctccgggg	ttcatgttgc	gcagaaccac	cagcacagt	8460
tatccgggtc	acttggaaa	tttgcatgt	agcttagaa	gaaatgcgt	gaagaactt	8520
gagacgcctt	tgtgacctcc	aagattttcc	atgcattcg	ccataatgtat	ggcaatgggc	8580
ccacggggcgg	cggctggc	gaagatattt	ctgggatcac	taacgtata	gttgttcc	8640
aggatgagat	cgtcataggc	cattttaca	aagcgcgggc	ggagggtgcc	agactgcgt	8700
ataatggttc	catccggccc	aggggcgtag	ttaccctcac	agatttgc	ttcccacgct	8760

ttgagttcag atgggggat catgtctacc tgcggggcga tgaagaaaac ggtttccggg	8820
gtaggggaga tcagctggg agaaagcagg ttcttagca gctgcgactt accgcagccg	8880
gtggggccctt aaatcacacc tattaccggc tgcaactggg agttaaagaga gctgcagctg	8940
ccgtcatccc tgagcagggg gccacttcg ttaagcatgt ccctgactcg catgtttcc	9000
ctgaccaaata ccgcagaag ggcgtcgccg cccagcgata gcagttcttga caaggaagca	9060
aagttttca acggttttag accgtccgccc gttaggtatgc tttagggcggt ttgaccaagc	9120
agttccaggc ggtccccacag ctccgttcacgg tgctctacgg catctcgatc cagcatatct	9180
cctcgtttcg cgggttgggg cggcttcgc tgtaacggcag tagtcgggtgc tcgtccagac	9240
gggccagggc catgttttc cacggggcga gggtcctcgatc cagcgtatgc tgggtcacgg	9300
tgaagggggtg cgctccgggc tgccgtcgcc ccagggtcgcc cttagggctg gtcctgtgg	9360
tgctgaagcg ctgcgggtct tgccctcgcc cgccggccat gtagcattt accatgggt	9420
catagtcacg cccctcccg gctggccct tgccgcgcag ctgcggctt gaggaggcgc	9480
cgacacgggg gcagtcaga ctttgagggg cgttagagctt gggcgcgaga aataccgatt	9540
ccggggagta ggcatcccgcc cccggggcc cgcagacggt ctgcattcc acgagccagg	9600
tgagctctgg cgggtcggtt tcaaaaaaccg ggtttcccccc atgctttt atgcgtttct	9660
tacctcttgtt ttccatgagc cgggttccac gtcgggtgc gaaaaggctg tccgtgtccc	9720
cgtatacaca cttgagaggc ctgtctcgatc gccgggttcc gccgtccctcc tcgtatagaa	9780
actcggacca ctctgagacg aaggctcgcc tccaggccag cacgaaggag gctaagtggg	9840
aggggttagcg gtcgttgcactt actagggggt ccactcgctc cagggtgtga agacacatgt	9900
cgccttcggc ggcatcaagg aaggtattt gtttataggt gtaggcacgc tgaccgggtg	9960
ttcctgaagg gggctataa aagggggtgg gggcgcgttc gtcctactc tcttcggcat	10020
cgctgtctgc gaggggccgc tgtaggggtt agtactccct ctcaaaaagcg ggcatgactt	10080
ctgcgctaag attgtcagtt tccaaaaacg aggaggattt gatattcacc tgcccgccgg	10140
tgtgcctt gagggtggcc gcttcatct ggtcaaaaa gacaatctt ttgtgtcaaa	10200
gcttgggtggc aaacgacccg tagagggcgatc tggacagcaa ctggcgatc gagcgcaggg	10260
tttgggtttt gtcgcgtatcg ggcgtcttcc tggccgcgtatc gtttagctgc acgtattcgc	10320
gcgcacacgca cggccattcg gaaaaagacgg tgggtgcgtc gtcgggactt aggtgcacgc	10380
gccaaccgcg gttgtgcagg gtgacaaggat caacgcgtt ggcgtacccctt ccgcgttaggc	10440
gctcggtggc ccagcagagg cggccgcctt tgccgcgcgca gaatggcggt agtgggtcta	10500
gctgcgtctc gtcggggggg tctgcgtccca cggtaaagac cccggggcagc aggcgcgcgt	10560
cgaagtagtc tatcttgcattt ctttgcgtt ctgcgcgtc cggggccaa	10620
gcgcgcgttc gtatgggtt agtgggggac cccatggcat ggggtgggtt agcgcggagg	10680
cgtacatgcc gcaaatgtcg taaacgtaga ggggtcttctt gagtattcca agatatgttag	10740
ggtagcatct tccaccgcgg atgctggcgc gcaacgtatc gtatagttcg tgcgaggag	10800
cgaggaggc gggaccggg ttgctacggg cgggtcgctc tgctcgaaag actatctgcc	10860
tgaagatggc atgtgagttt gatgatatgg ttggacgtcg gaaagacgtt aagctggcgt	10920
ctgtgagacc taccgcgtca cgcacgaagg aggcgttagga gtcgcgcgc ttgttgacca	10980
gctcgccgtt gacgtgcacgc tctaggccgc agtagtccag gttttccctt atgatgtcat	11040
acttatactt tccctttttt tccacagct cgcgggttag gacaaactct tcgcggctt	11100
tccagtagtc ttggatcgga aaccgtcggtt cctccgaacg gtaagagcctt agcatgtaga	11160
actgggttgc ggcctggtag ggcgcacatc cttttctac ggttagcgccg tattgcctgcg	11220
cggcccttcggc gagcggagggtg tgggtgagcg caaagggttc cctaaccatg actttggagg	11280
actgggtattt gaagtcaatgc tgctgcgtatc cgcggctc ccagagcaaa aagtccgtgc	11340
gctttttggc acgcgggtt ggcaggggcga aggtgacatc gttgaagagttt atctttcccg	11400
cgcgaggcat aaagggtcgatc gtgtatcgatc aggggtcccg cacctcgaa cgggtgtttaa	11460
ttacctgggc ggcgagcactt atctcgatcaaa agccgttgc gttgtggccc acaatgtaaa	11520
gttccaagaa ggcggggatc cccttgcgtt aaggcaattt tttaagttcc tcgttaggtga	11580
gctcttcagg ggagctgagc cgcgtctcg aaaggccca gtcgtcaaga tgagggttgg	11640
aagcgacgaa tgagctccac aggtcacggg ccattagcat ttgcagggtgg tcgcgaaagg	11700
tcctaaactg ggcacatcgatc gccatccctt ctgggtgtatc gcaatggaa gtaagcgggt	11760
cttgggtccca gcggtcccat ccaagggtcccg cggctaggatc tcgcgcggcg gtcactagag	11820
gctcatctcc gccgaacttc atgaccagca tgaaggccac gagctgcctt ccaaaaggccc	11880
ccatccaagt ataggctctt acatcgatgg tgacaaagag acgctcggtg cgaggatgcg	11940
agccgatcgg gaagaactgg atctccgcacc accagttggaa ggagtggctg ttgtatgtgg	12000
gaaagtagaa gtcctcgca cggggccgaac actcgatcgatc gcttttgcgtaa aaacgtgcgc	12060

agtactggca gcgggtgcacg ggctgtacat cctgcacgag gttgaccctga cgaccgcgca 12120
 caaggaagca gagtggaaat ttgagccccct cgcctggcggt gtttggctgg tggctttcta 12180
 ctteggctgc ttgtcccttga ccgtctggct gctcgagggg agttacggtg gatcgacca 12240
 ccaacggcccg cgagccccaaa gtccagatgt ccgcgcgcgg cggtcggagc ttgatgacaa 12300
 catcgccgag atgggagctg tccatggctt ggagctcccgg cggcgtcagg tcaggcggga 12360
 gcttcctgcag gtttacctcg catagccggg tcagggcgcg ggcttaggttc aggtgatacc 12420
 tgatttccag gggctgggtt gtggcggcgt cgatggctt caagaggccg catcccccgcg 12480
 ggcgcgactac ggtaaccgcgc ggcggggcggt gggccgcggg ggtgtccctt gatgatgcat 12540
 ctaaaaagcgg tgacgcgggc gggccccccgg aggttaggggg ggctcgggac cgcggggag 12600
 agggggcagg ggcacgtcgg cggcgcgcg gggcaggagc tggtgctgcg cgccggagggtt 12660
 gctggcgaac ggcacgcgcg ggcgggttat ctcctgaatc tggcgcctct gctgtgaagac 12720
 gacggggcccg gtgagcttga acctgaaaga gagttcgaca gaatcaattt cggtgtcggt 12780
 gacggcggcc tggcgaaaaa ttcctgcac gtctcctgag ttgtcttgcg aggcatctc 12840
 ggccatgaac tgctcgatct ctccttcctg gagatctccg cgtccggctc gctccacgg 12900
 ggcggcgagg tcgttggaga tgcgggcat gagctgcag aaggcggttga ggcctccctc 12960
 gttccagacg cggctgtaga ccacggccccc ttccgcacatc cggcgcgcga tgaccacctg 13020
 cgcgagattt agctccacgt gccggcgaaa gacggcgtat ttccgcaggc gctgaaagag 13080
 gttagttgagg gtggggcggt tgggttctgc cacgaagaag tacataaccc agcggcccaa 13140
 cgtggattcg ttgatatccc ccaaggccctc aaggcgctcc atggcctcgtt agaagtccac 13200
 ggcgaaggtt aaaaactggg agttgcgcgc cgacacgggtt aactcctccct ccagaagacg 13260
 gatgagctcg ggcacagtgtt cgcgcacccgc gctaccaaa gctacagggg cctcttc 13320
 ttcttcataatc tccttcataa taaggccctc cccttcctt tcttcggcg gcggtgggg 13380
 aggggggaca cggccgcgac gacggcgcac cgggaggcggt tcgacaaaagc gctcgatcat 13440
 ctccccccgcg cgacggcgca tggctcggtt gacggcgcgg ccttctcgc gggggcgcag 13500
 ttggaaagacg ccgcccgtca tggcccggtt atgggttggc ggggggctgc cgtgcggcag 13560
 ggatacggcg ctaacgatgc atctcaaaat ttgttgcgatc ggtactccgc caccgaggga 13620
 cctgagcgag tccgcacatc cccggatcgaa aaacctctcg agaaaggcgt ctaaccagtc 13680
 acagtcgaa ggtaggctga gcaccgtggc gggcggcage gggcggcggt cgggggttgc 13740
 tctggcggag gtgtgtcgatc tgatgtattt aaagttaggcg gtcttgcgatc ggcggatgtt 13800
 cgacagaagc accatgttcc tgggtccggc ctgtgtatc cgcaggcggtt cggccatgcc 13860
 ccaggcttcg ttttgcacatc ggcgcagggtc ttttgcgttgc tcttgcatgc gcctttctac 13920
 cggcacttct tcttctccctt cctttgttcc tgcacatctt tgcacatctatcg ctgcggcg 13980
 ggcggaggtt ggcggctctt tccctccatc cgtgtgacccc cgaagccct 14040
 catcggtca agcaggggcca gtcggcgcac aacgcgtcg gctaataatgg cctgtgcac 14100
 ctgcgtgagg gtagacttggaa agtcgtccat gtccacaaag cgtgtgtatc cggccgtgtt 14160
 gatgggttaa gtgcagttgg ccataacggc ccagttaaacg gtctgtgtac cggcgtcg 14220
 gagctcggtt tacctgagac gcgagtaagc ccttgcgttca aagacgtatc cggtgc 14280
 ccgcaccagg tactgggtatc ccaccaaaaaa gtgcggcgcg ggttgcgggtt agaggggcca 14340
 gcgtaggggtt gcccgggctc cggggggcgag gtcttccaaatc ataaggcgat gatatccgt 14400
 gatgtacccgt gacatccagg tggatgcggc ggcgggtgtt gaggcgcgcg gaaagtccacg 14460
 gacgcgggttc cagatgttgc gcaacggcggc aaagtgttcc atgggtcgggca cgctctggcc 14520
 ggtcaggcgc ggcacgtcgatc tggatgcgttca gaccgtgcac aaggagagcc tggtaaggcg 14580
 cactttccg tgggttgcgtt gataaatttcg caagggtatc atggcggacg accgggggttc 14640
 gaaccccccga tccggccgtc cggccgtatc catgcggta ccggccgggtt gtcgaacccca 14700
 ggtgtgcgac gtcagacaac gggggagcgc tccctttggc tcccttcag ggcggcggcga 14760
 tgctgcgtca gctttttgg ccactggccg cgcggcgtt aagcggttag gctggaaagc 14820
 gaaagcatta agtgggtcgcc tccctgtatc cggagggtt ttttccaaagg gttgagtcgc 14880
 gggggcccccgtt gttcgagttt cggggccggcc ggcactgcgc gaaacgggggtt tggctccccc 14940
 gtcatgcaccc accccgcctt ccaattccctt cggaaacagg gacgagcccc tttttgtttt 15000
 ttccctccatc catccgggtcc tggcggcgtt ggcggccctt cctcagcgcgc ggcaagagca 15060
 agagcagcgg cagacatgca gggcaccctc cccttcctt acccggtcag gaggggcaac 15120
 atccgcggct gacgcggcgg cagatgttgc ttacgaaccc cgcggcggcc ggcacccggca 15180
 ctacttggac ttggaggagg gcgaggccctt ggcggcgtt ggcggccctt ctcctgagcg 15240
 acaccccaagg gtgcagtcga agcgtgcacac ggcggaggcg tacgtgcgcgc ggcagaaccc 15300
 gtttcgcgac cgcggaggagg aggagcccgaa ggagatgcgg gatcgaaagt tccatgcagg 15360

gcgcgagttt	cggcatggcc	tgaaccgcga	gcgggttgctg	cgcgaggagg	actttgagcc	15420
cgacgcgcgg	accgggatta	gtcccgcgc	cgcacacgtg	gcggccgccc	acctggtaac	15480
cgcgtacgag	cagacggta	accaggagat	taacttcaa	aaaagctta	acaaccacgt	15540
gcmcacgctt	gtggcgcgc	aggaggtggc	tataggactg	atgcatctgt	gggactttgt	15600
aagcgcgtg	gagcaaaacc	caaatacgaa	gccgctcatg	gcgcagctgt	tccttatagt	15660
gcagcacacg	agggacaacg	aggcattca	ggatgcgcgt	ctaaacatag	tagagccgaa	15720
ggggcgctgg	ctgctcgatt	tgataaaacat	tctgcagagc	atagtggtgc	aggagcgcag	15780
cttgcgcctg	gctgacaagg	tggccgcct	taactattcc	atgctcagtc	tggcaagtt	15840
ttacgcccgc	aagatatacc	ataccctta	cgttccatata	gacaaggagg	taaagatcga	15900
ggggttctac	atgcgcattgg	cgctgaaggt	gcttacatttgc	agcgacgacc	tgggcgttta	15960
tcgcaacag	cgcacccaca	aggccgttag	cgtgagccgg	cggcgcgcgc	tcagcgaccg	16020
cgagctgatg	cacagcctgc	aaagggccct	ggctggcacc	ggcagcggcg	atagagaggc	16080
cgagtcctac	tttgcgcgg	gctgcgtaccc	gctgcgtggcc	ccaagccgac	gcgccttgaa	16140
ggcagctggg	gccggacctg	ggctggcggt	ggcaccccg	cgcgcgtggca	acgtcgccgg	16200
cgtggaggaa	tatgacgagg	acgtatggat	cgagccagag	gacggcgagt	actaagcggt	16260
gatgtttctg	atcagatgt	gcaagacgca	acggacccgg	cgtgcgggc	ggcgcgtcag	16320
agccagccgt	ccggccttaa	ctccacggac	gactggcgc	agtcatgg	ccgcacatcatg	16380
tcgcgtactg	cgcgcacacc	tgacgcgtt	cggcagcagc	cgcaggccaa	ccggctctcc	16440
gcaattctgg	aaggcgttgtt	ccccggcgc	gcaaaaaacc	cgcacgagaa	ggtgctggcg	16500
atcgtaaacg	cgctggccga	aaacaggccc	atccggcccg	atgaggccgg	cctggcttac	16560
gacgcgcgtc	ttcagcgcgt	ggctcggttac	aacagcagca	acgtgcagac	caacctggac	16620
cggtgggtgg	gggatgtgcg	cgaggccgt	gctgcgtgt	agcgcgcgc	gcagcaggggc	16680
aacctggct	ccatggttgc	actaaacgccc	ttccctgat	cacagccgc	caacgtgcgc	16740
cggggacagg	aggactacac	caacttgc	agcgcactgc	ggctaattgtt	gactgagaca	16800
ccgcaaatg	aggtgttatca	gtccgggcca	gactattttt	tccagaccag	tagacaaggc	16860
ctgcagaccc	taaacctgag	ccaggctttc	aagaacttgc	aggggctgt	gggggtgcgg	16920
gctcccacag	gcgaccgcgc	gaccgtgtct	agcttgcgt	cgcacaactc	gcgcctgttgc	16980
ctgctctaa	tagccctt	cacggacagt	ggcagcgtgt	ccccggacac	ataccttaggt	17040
cacttgcgt	cactgtacc	cgaggccata	ggtcaggcgc	atgtggacga	gcataacttcc	17100
caggagatta	caagtgttag	ccgcgcgt	gggcaggagg	acacgggcag	cctggaggca	17160
accctgaact	acctgcgtac	caacccggcgg	caaaaaatcc	cctcgttgc	cagtttaaac	17220
agcgaggagg	agcgcatttt	gcttatgt	cagcagacgc	tgagccttaa	cctgtatgcgc	17280
gacggggtaa	cgcccaacgt	ggcgcgtggac	atgaccgcgc	gcaacatgg	accgggcac	17340
tatgcctcaa	accggccgtt	tatcaatgc	ctaattggact	acttgcac	cgccggccgc	17400
gtgaaccctg	agtatttac	caatgcac	ttgaacccgc	actggctacc	gccccctgg	17460
ttctacaccc	ggggatttgc	ggtgccccag	ggtaacgat	gattcctct	ggacgacata	17520
gacgacagcg	tgtttcccc	gcaaccgcag	accctgt	agttgcaaca	acgcgagcag	17580
gcagaggccg	cgctgcgaaa	ggaaagctc	cgcagccaa	gcagcttgc	cgatctaggc	17640
gctgcggccc	cgcggtcaga	tgcttagtgc	ccatttccaa	gcttgatagg	gtctcttacc	17700
agcaactcgc	ccacccgc	gcccgcgt	ggcgaggagg	agtagctaaa	caactcgcgt	17760
ctgcagccgc	agcgcgaaaa	gaacctgc	ccggcgtt	ccaaacaacgg	gatagagagc	17820
ctagtggaca	agatgatgt	atggaagacg	tatgcgcagg	agcacaggga	tgtgcccggc	17880
ccgegcggcc	ccacccgtc	tcaaaggcac	gaccgtc	ggggcttgc	gtgggaggac	17940
gatgactcgg	cagacgacag	cagcgtct	gatttggag	ggagtggca	cccggttgc	18000
caccttcgc	ccaggctgg	gagaatgtt	aaaaaaaaa	catgtatgc	aataaaaaaac	18060
tcaccaaggc	catggcacc	agcgttgtt	ttcttgcatt	cccttgcata	tgcggcgcgc	18120
ggcgatgtat	gaggaagg	ctccctcc	ctacgagac	gttgcgc	cgccgcacgt	18180
ggcggcggcg	ctgggttac	ccttcgtat	ccccctggac	ccgcgcgttgc	tgccctccgc	18240
gtacctgcgg	cctaccgggg	ggagaaacag	catccgttac	tgcgttgttgc	cacccctatt	18300
cgacaccacc	cgtgtgtacc	tttgtggacaa	caagtcaacg	gttgtggcat	ccctgaacta	18360
ccagaacac	cacaccaact	ttctaacc	ggtcatca	aacaatgact	acagccggg	18420
ggaggcaagc	acacagacca	tcaatcttgc	cgaccggc	cactggggc	gacgcac	18480
aaccatctg	cataccaaca	tgccaaatgt	gaacgat	atgtttacca	ataagttaa	18540
ggcgccgggt	atgggtgc	gctcgcttac	taaggacaaa	cagggtggac	tgaaatacga	18600
gtgggtggag	ttcagcgtc	ccgaggccaa	ctactcc	accatgacca	tagaccttat	18660

gaacaacgcg atcgtggagc actacttcaa agtggcagg cagaacgggg ttctggaaag 18720
 cgacatcggtt gtaaagtttt acaccgcga cttcaactg gggtttgacc cagtcactgg 18780
 tcttgcata cctggggtat atacaaacga agcctccat ccagacatca ttttgctgcc 18840
 agatgcggg gtggacttca cccacagccg cctgagcaac ttgttggca tccgcaagcg 18900
 gcaacccttc caggagggtt ttaggatcac ctacgatgac ctggagggtt gtaacattcc 18960
 cgcactttt gatgtggacg cttaccaggc aagctgaaa gatgacaccc aacaggcgg 19020
 gggtggcga ggcggcggca acaacagtgg cagcggcgcg gaagagaact ccaacgcggc 19080
 agctgcggca atgcagccgg tggaggacat gaacgatcat gccattcgcg gcgcacaccc 19140
 tgccacacgg gcccggaggaga agcgcgctga ggccgaggca gcccggcaag ctggccccc 19200
 cgctgcggag gctgcacaac ccgaggctga gaagctctag aaaaaaccgg tgattaaacc 19260
 cctgacagag gacagcaaga aacgcagta caacctaata agcaatgaca gcacccctac 19320
 ccagttcccg agctgttacc ttgcataccaa ctacggcgc cctcaggccg ggatccgctc 19380
 atggaccctt ctttgcactc ctgacgttaac ctgcgcgtcg gagcaggat actggtcgtt 19440
 gccccacatg atgcacggacc cctgtacccctt ccgcctccacg cggcagatca gcaacttcc 19500
 ggtgggtggc gccgagctgt tgccctgtca ctccaagagc ttctacaacg accaggccgt 19560
 ctactcccg ctcatccggc agtttacccctc tctgaccaccc gtgttcaatc gtttcccg 19620
 gaaccagatt ttggcgcggc cggccagcccc caccatcacc accgtcagtg aaaacgttcc 19680
 tgctctcaca gatcacggga cgctaccgtc ggcacacgc atccggaggag tccagcgt 19740
 gaccattact gacgcccagac gccgcacccctg cccctacgtt tacaaggccc tgggcataatg 19800
 ctcgcgcgc gtccttatcga gccgcacttt ttgagcaagc atgtccatcc ttatatcgcc 19860
 cagcaataac acaggctggg gctgtcgctt cccaagcaag atggttggcg gggccaagaa 19920
 ggcgcgcgc caacaccccg tgcgcgtcg cgggcactac cggcgcgcgc ggggcgcgc 19980
 caaacgcggc cgcaactggc gcaccacccgt cgatgacgcg atcgcacgcgg tgggtggagga 20040
 ggcgcgcac tacacgcggc cggccgcgc agtgtccacc tggacgcgcgg ccattcagac 20100
 cgtgggtgcgc ggagcccgcc gctacgctaa atatggaga cggcggaggc gcgttagcgc 20160
 tcgcgcgcgc cgccgaccccg gcaactggcgc ccaacgcgcg gcccggccgc tgcttaaccg 20220
 cgcacgtcgc accggccgac gggcgccat gcgaggccgt cgaaggctgg cccgcgggtat 20280
 tgtcaactgtg ccccccaggt ccaggcgacg agcggccgc gcaagcgcgc cggccattag 20340
 tgctatgact cagggtcgca gggcaacgt gtactgggtg cggcactcgg ttagcgcct 20400
 ggcgcgtgccc gtgcgcaccc gcccccccgca caactagatt gcaataaaaaa actacttaga 20460
 ctcgtactgt tgtatgttac cagcggcggc ggccgcgcgc gaagctatgt ccaagcgc 20520
 aatcaaagaa gagatgttcc aggtcatcgc gccggagatc tatggccccc cgaagaagga 20580
 agagcaggat tacaagcccc gaaagctaaa gcccgtcaaa aaaaaaaaaga aagatgtga 20640
 tgcgtatgaa cttgacgacg aggtggaaact gttgcacgcg accgcgcggc ggcgcacgg 20700
 acagtggaaa ggtgcacgcg taagacgtgt ttgcgcaccc ggacccaccc tagtctttac 20760
 gcccgggtgag cgctccaccc gcacctacaa ggcgcgttat gatgagggtgt acggcgcacga 20820
 ggacctgtt gaggcggca acgagcgcct cggggagttt gcttacggaa agcggcataa 20880
 ggacatgtgt gcgttgcgcg tggacgaggg caacccaaaca ccttagcctaa agcccggtac 20940
 actgcagcag gtgcgtcccg cgcttgcacc gtccgaagaa aagcgcggcc taaagcgcga 21000
 gtctgggtgac ttggcaccac ccgtgcagct gatggtaccc aagcgtcgc gactggaaaga 21060
 tgtcttggaa aaaatgaccg tggagcctgg gctggagccc gaggtcccgcg tgccggcaat 21120
 caaggcagggtg gcaccgggac tggcgcgtca gaccgtggac gtcagatacc caccaccc 21180
 tagcactagt attgcccactg ccacagaggg catggagaca caaacgtccc cgggtgcctc 21240
 ggcgggtggca gatggcgcgg tgcaggcggc cgctgcggcc gcgttcaaga cctctacgga 21300
 ggtgcacaaacg gaccggcgtt gttttcgtgt ttcaagcccc cggcgtccgc gcccgttcaag 21360
 gaagtacgcgc gcccggcagcg cgctactgcgc cgaatatgcctt ccatcgcgc 21420
 taccggcgtt tgcgtggct acacccatccg ccccaagaaga cgagcaacta cccgcaccc 21480
 aaccaccaact ggaacccggc gcccggcgtcg cggcgcgcac cccgtgtccgg ccccgatttc 21540
 cgtgcgcagg gtggcgtcg aaggaggcag gaccctgggtg ctgcacccacag cggcgttacca 21600
 ccccaacatc gttttaaaggc cggctttgt ggttcttgc gatatggccc tcacccgtccg 21660
 cctccgtttc cccggcggg gatcccgagg aagaatgcac cgtaggaggc gcatggccgg 21720
 ccacggccctg acggggcggca tgcgtcgtgc gcaccacccgg cggcggcgcgc cgtcgcaccc 21780
 tgcgtatgcgc ggcggatcc tggcccttccct tattccactg atgcgcgcgg cgattggcgc 21840
 cgtggggaa attgcatccg tggcccttgc ggcgcagaga cactgattaa aaacaagttt 21900
 catgtggaaa aatcaaataaa aatgttgcgatcc ctctcactt cgcgttgcgc tgtaactatt 21960

ttgtagaatg gaagacatca actttgcgtc actggccccg cgacacggct cgcgccgtt 22020
 catggaaac tggcaagata tcggcaccag caatatgagc ggtggcgct tcagctggg 22080
 ctcgctgtgg agcggcatta aaaattcgg ttccgcgtt aagaactatg gcagcaaagc 22140
 ctggaacagc agcacaggcc agatgctgag ggacaagtt aaagagcaa atttccaaca 22200
 aaaggtggta gatggcctgg cctctggcat tagcgggtg gtggacctgg ccaaccaggc 22260
 agtgc当地 aagattaaca gtaagcttga tccccgcctt cccgttagagg agcctccacc 22320
 ggccgtggag acagtgtctc cagagggcg tggcggaaaag cgccgcgcac ccgacaggga 22380
 agaaaactctg gtgacgcaaa tagacgagcc tccctcgat gaggaggcac taaagcaagg 22440
 cctgcccacc acccgtccca tcgcgcctt ggctaccggg gtgctggcc agcacacacc 22500
 cgtaacgcgtg gacgcgcctc ccccccgcga caccggcag aaacctgtgc tgccaggccc 22560
 gtccgcgtt gttgtaaaaacc gtccttagcc cgccgcgcctt ccagcggtcc 22620
 gcgatcgatg cggcccgtag ccagtggcaaa ctggcaaggc acactgaaca gcatcgatgg 22680
 tttgggggtt caatccctga agcggcgacg atgcttgcgt tagttaacgt gtcgtatgt 22740
 tgtcatgtat ggcgtccatgt cggccggcaga ggagctgtg agccggcgcc cgccgcgtt 22800
 ccaagatggc taccgcctcg atgatgcgcg agtggctta catgcacatc tcggccagg 22860
 acgcctcgga gtacgtggc cccggggctgg tgcgttcgc ccgcgcacc gagacgtact 22920
 tcagcctgaa taacaagtt agaaacccca cggtggccgc taacgcacgac gtgaccacag 22980
 accggctcta ggcgttgacg ctgcgggttca tccccgttga ccggcaggat actgcgtact 23040
 cgtacaaggc gcgggttccacc cttagctgtgg gtgataaccg tggcttagac atggcttcca 23100
 cgtactttga catccggggc gtgctggaca gggggccctac ttttaagccc tactctgca 23160
 ctgcctacaa cgcactggcc cccaaagggtt cccccaactc gtgcgagttt gaacaaaatg 23220
 aaactgcaca agtggatgtt caagaactt acgaagagga gaatgaagcc aatgaagctc 23280
 aggccgcgaga acaggaacaa gctaagaaaa cccatgtata tgcccaggct ccactgtccg 23340
 gaataaaaaat aactaaagaa ggtctacaaa taggaactgc cgacgcccaca gtagcagggt 23400
 ccggcaaaa aatttcgca gacaaaactt ttcaacctga accacaagta ggagaatctc 23460
 aatggaaacga agcggatgcc acagcagctg gtggagggt tctttttttt acaactccca 23520
 taaaacccctg ctatggctca tacgttagac ccaccaattc caacggcgga cagggcggtt 23580
 tgggtgaaca aaatggtaaa ttggaaagtc aagtcgaaat gcaattttttt tccacatcca 23640
 caaatgcccc aatgaagtt aacaatatac aaccaacagt tttttttttt agcgaagatg 23700
 taaaacatgaa aactccagat actcatctt cttttttttt taaaatgggg gataaaaaatg 23760
 cccaaatgtt gcttggacaa caagcaatgc caaacagacc aaattttttt gctttttttt 23820
 acaattttttt tgggtctcatg tattacaaca gcacaggtaa catgggtgtc cttgtgtgtc 23880
 aggcatcgca gttgaacgct gtttagatt tgcaagacag aaacacacagag ctgtcctacc 23940
 agctttgtt tttttttttt ggcgcacagaa caagataactt ttcaatgtt gatcaagctg 24000
 ttgacagcta tttttttttt gtcagaattt ttgagaacca tggacttgg gatgagttgc 24060
 caaattttttt tttttttttt ggtggattt ggattttttt tttttttttt gctgtttttt 24120
 caactgtgtc taacggggac caaggcaata ctacctggca aaaaatgtt acattttttt 24180
 aacgcaatga aataggggtt gggaaataact ttgcccattttt aatattttttt aatgccaacc 24240
 tatggagaaaa tttttttttt tccaaatattt cgctgtttttt gccagacaag cttttttttt 24300
 accccccaaa tttttttttt tttttttttt tttttttttt tttttttttt aacaaggcgg 24360
 tgggtggctcc tttttttttt gttttttttt gttttttttt gttttttttt tttttttttt 24420
 acatggacaa ctgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt 24480
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 24540
 ccattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 24600
 atgttaacat gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 24660
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 24720
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 24780
 ccacgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 24840
 catcgccaa ctggggcagca tttttttttt tttttttttt tttttttttt tttttttttt 24900
 cccctttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 24960
 ttgacggaaac tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 25020
 ctgttagctg gcccggcaac gaccgcgtt tttttttttt tttttttttt tttttttttt 25080
 cagttgacgg ggagggtttt aacgtgtttt agtgcacat gacaaaggac tttttttttt 25140
 tgcagatgtt ggccaactac aatattttttt accagggtttt tttttttttt tttttttttt 25200
 aagaccggcat gtactcgat tttttttttt tttttttttt tttttttttt tttttttttt 25260

atactaaata caaagattat cagcagggtt gaatttatcca ccagcataac aactcaggct 25320
 tcgttaggcta cctcgctccc accatgcgcg agggacaagc ttaccccgct aatgtccct 25380
 acccaactaat aggcaaaaacc gcgggttgcata gtattaccca gaaaaagttt ctttgcacc 25440
 gcacccctgtg gcgcatcccc ttctccagta actttatgtc catgggtgcg ctcacagacc 25500
 tgggccaaaaa ccttctctac gcaaactccg cccacgcgt agacatgacc tttgaggtgg 25560
 atcccatgga cgagccccacc cttctttatg ttttgttga agtctttgac gtggccgtg 25620
 tgcaccagcc gcaccgcggc gtcatcgaga ccgtgtaccc ggcacgcggc ttctcgccg 25680
 gcaacgcac aacataaaaga agcaagcaac atcaacaaca gctgccgcga tgggctccag 25740
 tgagcaggaa ctgaaagcca ttgtcaaaga tcttgggtt gggccatatt ttttggcac 25800
 ctatgacaag cgcttcccag gctttgttcc cccacacaag ctcgcctgcg ccatagttaa 25860
 cacggccggc cgcgagactg ggggcgtaca ctggatggcc tttgccttggaa acccgcgctc 25920
 aaaaacatgc tacctcttg agccctttgg ctttctgtc caacgtctca agcagggtta 25980
 ccagtttag tacgagtac tcttgcgcgc tagccatt gccttctcc cgcaccgcgt 26040
 tataacgcgt gaaaagtcca cccaaagcgt gcaggggccc aactcgccgcg cctgtggcct 26100
 attctgctgc atgtttctcc acgccttgc caactggccc caaactccca tggatcacaa 26160
 ccccaccatg aaccttatta cccgggttacc caactccatg cttAACAGTC cccagggtaca 26220
 gcccaccctg cgccgcacc accaagcgt ctacagettc ctggagcgcc actcgcccta 26280
 ctteccgcagc cacagtgcgc aaatttaggag cggccacttct ttttgtcaact tgaaaacat 26340
 gtaaaaataa tgtacttagga gacactttca ataaaaggca atgaaaaatttt ttgtacactc 26400
 tcgggtgatt atttacccccc acccttgcgcg tctgcgcgt ttaaaaatca aagggggtct 26460
 gccgcgcac gctatgcgc actggcaggg acacgttgcg atactgggtt ttagtgcctc 26520
 acttaaaactc aggccacaacc atccgcggca gctcgggtaa gttttcaactc cacaggctc 26580
 gcaccatcac caacgcgtt agcagggtgg ggcgcgatata cttaagatcg cagttgggc 26640
 ctccgcctg cgccgcgcgag ttgcataca cagggttaca gcactggaaac actatcagcg 26700
 cccgggtggc cacgtggcc accacgtctc tgcggagat cagatccgcg tccaggctc 26760
 cccgcgttgc cagggcgaac ggagtcaact ttggtagctg ctttcccaaa aagggtgcac 26820
 gcccaggcgtt tgagggtcac tcgcaccgtt gtggcatcg aaggtgaccg tgcccagtct 26880
 gggcgtttagg atacagcgcc tgcataaagg ccttgcgtctt cttAAAGCC acctgaggct 26940
 ttgcgccttc agagaagaac atgcgcgaag acttgcgcga aaactgattt gccggacagg 27000
 cccgcgtcatg cacgcagcac cttgcgtcg tttggagat ctgcaccaca ttcggccccc 27060
 accgggttctt cacgatcttgc ctttgcgtactt actgcgttcc cagcgccgcg tgcccgttt 27120
 cgcgcgtcac atccatttca atcacgtgtc ctttattat cataatgtc cccgttagac 27180
 acttaagctc gccttcgatc tcaagcgccgc ggtgcagcca caacgcgcag cccgtggct 27240
 cgtgggtctt gtaggttacc tctgcataacg actgcaggta cggccgtcagg aatcgcccc 27300
 tcatcggtcac aaagggtttt tgctgtgtt ggtcagctg caacccgcgg tgctcctcg 27360
 tttagccagttt cttgcatacg gccgcgcagag cttccacttgc gtcaggcagt agcttgaagt 27420
 ttgcctttag atcgttatcc acgtggactt tgcgttactt cttgcatacg cgcgcgcgcgac 27480
 ccttctccca cgcagacacg atcggcaggc tcaagcggtt tattcaccgtt ctttacttt 27540
 cccgcgttact ggacttccctt gcatccgcatt accccgcgcgactgggtcg 27600
 cttcatttgc cccgcgcacc gtgcgttac cttcccttgc gtgttgcattt agcaccgggtg 27660
 gggttgcgtt accccaccatt tgcgttgcgcgc catcttctctt ttcttgcgttgc 27720
 tcacctctgg ggatggcggc cgtcgccgtt tggagaggg ggcgttctt ttctttttgg 27780
 acgcaatggc caaatccgcgc gtcagggtcg atggccgcgg gtcgggtgtg cggccacca 27840
 ggcgcacactg tgacgacttct tcttcgttcc cggactcgag acgcgcctc agccgccttt 27900
 ttggggggcgc gcccggaggc ggcggcgac ggcgcggggc cggagacgtcc tccatgggtt 27960
 gtggacgtcg cgcgcaccgc cgtccgcgtt cgggggttgc ttgcgtcg tcccttccc 28020
 gactggccat ttcccttcc tataaggcaga aaaagatcat ggagtcaactt gagaaggagg 28080
 acaggcttaac cggcccttgc ggttgcgcgc cccacgcctc caccgttgc gccaacgcgc 28140
 ctaccaccc cccgcgtcgag gcaaaaaaccgc ttgaggagga ggaagtgtt atcgaggcagg 28200
 acccagggtt tgtaagcgaa gacgcacgaaatcgctcaactt accaacagag gataaaaaagc 28260
 aagaccagga cgacgcacgag gcaaaacggagg aacaaggctgg ggggggggac caaaggcatg 28320
 ggcgacttactt agatgtggaa gacgcacgtgc tggtaagca tctgcacgcg cagtgcgc 28380
 ttatctgcga cgcgttgcga gacgcacgcg atgtgcctt cggccatagcg gatgtcagcc 28440
 ttgcctacga acgcacactg ttctcaccgc gctgtacccca caaacgccaa gaaaacggca 28500
 catgcgagcc caacccggcgc ctcaacttctt accccgttatt tgccgtgcca gaggtgttt 28560

ccacctataca catcttttc caaaaactgca agataccct atcctgccgt gccaaccgca 28620
 gccgagcggca aagcagctg gccttgcggc agggcgctgt cataacctgat atcgccctcg 28680
 tcgacgaagt gccaaaaatc ttggagggtc ttggacgcga cgagaagcgc gcccggaaacg 28740
 ctctgcaca aaaaaacagc gaaaatgaaa gtcactgtgg agtgctgggtg gaacttgagg 28800
 gtgacaacgc ggccttagcc gtgctgaaac gcagcatcga ggtcacccac tttgcctacc 28860
 cggcacttaa cttccccccc aagggttatga gcacagtcat gagcgagctg atcgtgcgc 28920
 gtgcacgacc cttggagagg gatgcaaact tgcaagaaca aaccggaggag ggcctacccg 28980
 cagttggcga tgagcagctg ggcgcgtggc ttgagacgcg cgcgcctgcg gacttggagg 29040
 agcgacgcaa gctaattatgt gcccgcgtgc ttgttaccgt ggagctttag tgcatgcgc 29100
 ggttcttgc tgaccggag atgcagcgcgca agctagagga aacgcttgcac tacaccctt 29160
 gccaggggcta cgtgcgcacg gcctgcacaa tttccaaacgt ggagctctgc aacctggct 29220
 cttacccttgg aatttgcac gaaaaccgcg ttggggcaaaa cgtgcttcat tccacgccta 29280
 agggcggaggc ggcgcgcgc acgtccgcg actgcgttta cttattttcg tgctacacct 29340
 ggcacaaacggc catgggcgtg tggcagcgt gcctggagga ggcacacccg aaggagctgc 29400
 agaagctgct aaagcaaaac ttgaaggacc tatggacggc cttcaacgcg cgctccgtgg 29460
 ccgcgcaccc ggcggacatt atctccccc aacgcctgtc taaaaccctg caacagggtc 29520
 tgccagactt caccatcaa agcatgttc aaaaacttag gaactttctg ctagagcgtt 29580
 caggaattct gcccgcacc tgctgtgcgc ttccctagcga ctttgtgccc attaagtacc 29640
 gtgaatgccc tccggcgtt tggggtcaact gctaccccttgcagcttgcaccc aactacc 29700
 cttaccactc cgacatcatg gaagacgtga gcccgtgacgg cctactggag tgcactgtc 29760
 gctgcaaccc atgcaccccg caccgcgtccc tggctctgca ttccacaactg cttagcga 29820
 gtcaaaattt cggtagctt gagctgcagg gtccttcgcg tgacgaaaag tccgcggctc 29880
 cggggttgaa actcactccg gggctgtgga cgtcggttta ctttcgcacaa tttgtac 29940
 aggactacca cgcacccacgat attaggttct acgaagacca atcccgcccc ccaaattgcgg 30000
 agcttaccgc ctgcgttattt accccaggggcc acatccctgg ccaatttgcac gccattaaaca 30060
 aagcccgcca agagttctg ctacgaaagg gacgggggggt ttacttggac ccccgatccg 30120
 gcgaggagct caacccaaatc ccccccgcgc cgcagcccttgcacgcggcc cggggcccttgc 30180
 cttcccgagga tggcacccaa aaaaagctg cagctgcgcgc cggccgcacc cacggacgag 30240
 gaggaaatact gggacagtca ggcagaggag gttttggacg aggaggagga gatgatggaa 30300
 gactgggaca gcctagacga ggaagcttcc gaggccgaag aggtgtcaga cgaaacaccg 30360
 tcacccctcg tcgcattttcc ctcgcggccg ccccaaaaaat cgcaaccgt tcccgatttgc 30420
 gctacaacctt ccgccttcgc ggcgcgcgc gcaactgcgc ttcgcgcacc caaccgttgc 30480
 tgggacacca ctggaaaccag ggcggtaag tctaaggcgc cggccgcgtt agcccaagag 30540
 caacaacacgc gccaaggcata cgcgtcggtt cgcgtgcaca agaacgcctat agttgttgc 30600
 ttgcacagact gttggggcaatcttc gcccggcgt ttcttcgttca ccatcaccgc 30660
 gtggccttc cccgttacat ctcgttatttac taccgttcatc tctacagccc ctactgcacc 30720
 ggcggcggcgc gcaacacgc cagcggccac gcagaagcaa aggcgcaccgg atagcaagac 30780
 tctgacaaatcccaaaaaat ccacagcggc ggcagcggc gggaggaggag cactgcgtct 30840
 ggcggccaaac gaacccgtat cggccggcga gcttagaaac agatttttc ccactctgt 30900
 tgctatattt caacagagca gggggcaaga acaagagctg aaaataaaaaa acaggtctct 30960
 ggcgccttc accccgcagct gctgttatca caaaaagcga gatcgttgc ggcgcacgct 31020
 ggaagacgcg gaggctcttct tcagcaataa ctgcgcgtt actctttaagg actagttcg 31080
 cggcccttcttcaaaatataag cggcggaaact acgtcatctc cagcggccac accccggcc 31140
 agcacctgtc gtcagcgcata ttatgagcaa gggaaatccc acggccctaca tggggaggta 31200
 ccagccacaa atgggacttg cggctggagc tgcccaagac tactcaaccc gaataaaacta 31260
 catgagcgcg ggaccccaata tgatatcccc ggtcaacggaa atcccgcccc accgaaaccg 31320
 aatttccttc gaacaggccg ctattaccac cacacccgtt aataaccctt atccccgtt 31380
 ttggcccgct gcccgggtt accagggaaag tcccgccccc accactgtgg tacttcccag 31440
 agacgcccac gcccggcgtt agatgactaa ctcaggggcg cagcttgcgg gcccgttcc 31500
 tcacagggtt cggcgtcccg ggcagggtt aactccacccgaaaatcagag ggcggaggat 31560
 tcagctcaac gacgagtcgg tgagctccctc tcttggcttc cgtccggacg ggacattca 31620
 gatcggccgc gctggccgtt cttcatttac gcccggctcag ggcgttccaa ctctgcagac 31680
 ctcgtccctcg gacccggcgtt ccggaggcat tggaactcta caatttttgg aggagttcg 31740
 gccttcgtt tactcaacc cttttctgg acctcccgcc cactaccggg accagttat 31800
 tcccaactttt gacgcggtaa aagactcggc ggacggctac gactgaatga ccagtggaga 31860

ggcagagcaa ctgcgcctga cacacctcgaa ccactgcccgc cgccacaagt gctttggcccg 31920
 cggctccggat gagtttgtt actttgaatt gccccaaagag catatcgagg gccccggcga 31980
 cggcgtcggc ctcaccaccc agtagagact tacacgtgc ctgattcggg agtttaccaa 32040
 gcggccctgt ctatggagc gggagcgggg tccctgtgtt ctgaccgtgg tttgcaactg 32100
 tcctaaccct ggattacatc aagatcttat tccatcaac taacaataaa cacacaataa 32160
 attacttaaaaatcagt cagcaaactt ttgtccagct tattcagcat cacctccctt 32220
 ccctcctccc aactctggta tttcagcagc cttttagctg cgaactttctt ccaaagtctt 32280
 aatggggatgt caaatcttc atgttcttgc ctctccgcac ccactatctt catattgtt 32340
 cagatgaaac gcgcaggacc gtctgaagac accttcaacc ctgtgttaccc atatgacacg 32400
 gaaaccggcc ctccaaactgt gccttcctt accccctccct ttgtgtcgcc aaatgggttc 32460
 caagaaaatc cccccggagt gctttctttg cgtcttcag aacctttgtt tacctcacac 32520
 ggcatgcttgc cgctaaaaat gggcagcggc ctgtccctgg atcaggcagg caaccttaca 32580
 tcaaaatacaatc tcaactgtttc tcaaccgctt aaaaaaaaaaactt aatccaatat aactttggaa 32640
 acatccgcgc cccttacatc cagctcaggc gcccataacc tggccacaaac ttgcctttg 32700
 gtggctcttg acaacactctt taccatcgaa tcacaagcac cgcttacccgt gcaagactca 32760
 aaaaatcttgc ttgcatttttca aagggccactt acagtgttag atggaaaactt ggccttcag 32820
 acatcagcccc ccctctctgc cactgataac aacgcctca ctatcactgc ctcaccctt 32880
 cttaactactg caaatggtag tctggctgtt accatggaaa accccacttta caacaacaat 32940
 gggaaaacttgc ggctaaaat tggcggtctt ttgcagggtt ccaccgactc acatgcacta 33000
 acacttaggtt ctggcagggtt ggttgcaggta cataacaattt tgctacatac aaaagttaca 33060
 ggcgcataatgggtt ggttgcataatc atctggcaac atggactta aacttggaga tggccttat 33120
 gtggatagcg cgggtcctaa cccaaaaactt cataattatc taaataccac aaaaggccctt 33180
 gcttttgcaca acaccgcaat aacaattaaac gctggaaaag ggttggattt taaaacagac 33240
 tcctcaaaacg gaaatcccat aaaaaacaaaaaatttgcatac gcttacataa taataccaaat 33300
 ggagctatgg ttgcaaaaactt tggaaacagggc ctcttttttgc acagctccgg agccataaca 33360
 atgggcagca taaacaatga cagacttactt ctggacaaacc caccagaccc atccccaaat 33420
 tgcagaatttgc ttccagatcaa agactgcaggc ctaactctgg cgcttacaaaatgtggcagg 33480
 caaattttttgcactgtttc agctttggca gtatcaggta atatggccttc catcaatgg 33540
 actctaagca gtgttactt ggttcttgc ttgcatttttgc acggagggtt tatgttcaat 33600
 tcatcactgg acaaaacagta ttggaaactttt agaaacgggg actccactaa cggtcaacca 33660
 tacactttagt ctgttgggtt tatgcacaaactt caaaaatgtt accccaaaaac tcaaaatggaa 33720
 actgcacaaaaa gtaatatttgtt tagccagggtt tatcttaatgc tgacacatgc taaaccattt 33780
 cattttacttta ttagctaaa tggaaacagat gaaaccaacc aagtaagccaa atactcaata 33840
 tcatttcaggat ggtcctggaa cagttggacaa tacactatgc acaaatttgc caccattcc 33900
 tataccttctt cctacattgc ccaggaaataa agaatcgta acctgttgc tggttatgtt 33960
 caacgtgtttt attttcaat tgcaggaaaat ttcaagtcat ttttcatca gtagtataatc 34020
 cccaccacca catacgcttacttacatcacc gtaccttataatcactcaca gaacccttagt 34080
 attcaacccgtt ccaccctccctt cccaaacacac agagtacaca gtcccttctc cccggctggc 34140
 cttaaacacgc atcatatcat gggtaacaga catatttttgc ggttgcataatccatcacc 34200
 ctccctgtcgac gccaaacgcgtt catcaggat gttataaaac tccccgggca gctcgctt 34260
 gttcatgtcg ctgttccaggct gctgagccac aggctgtgtt ccaacttgcgtt gttgctcaac 34320
 gggcggcggaa ggagaagtcc acgccttacat ggggttagag tccataatcgat gcatcaggat 34380
 agggcgggtgg tgctgcaggca ggcggcgaat aaactgtgtc cggccggcgtt ccgtcctgc 34440
 ggaataacaac atggcagggtt ttccttcaggc gatgattcgc accggccgcga gcttacggcg 34500
 ctttgccttc cgggcacacgc agcgcacccctt gatctactt aagtcttgcac agtaactgc 34560
 gcacagtacc acaatatttgtt taaaatccc acagtgcaggc ggcgtgttacatc caaagctcat 34620
 ggcggggacc acagaacccca cttggccatc ataccacaaag cgcaggtaga ttaagtggcg 34680
 accccctcata aacacgttgc acataaaacat taccttttttgc gcatgttgc ttttccaccac 34740
 ctccctgttacatcataaaacc tctgatttttgc tccaccacca tccataacca 34800
 gctggccaaa acctggccgc cggctatgcgtt ctgcaggaa cccggactgg aacaatgaca 34860
 gtggagagcc caggactcgat aaccatggat catcatgttgc gtcgtatgc ttttccaccac 34920
 acaacacacagg cacacgttgc acatacttccctt caggattaca agtcttgcac gcttacggcg 34980
 catatcccaccc ggaacaaccc accctgttgcacat cagcttacat cccacactgc agggaaagacc 35040
 tcgcacgttgc acacgttgc ttttccaccacca ttttccaccacca 35100
 ctccctgttacatcataaaacc ttttccaccacca 35160

agtgcgccga gacaaccgag atcgtttgg tcgttagtgtc atgccaaatg gaacgcccga 35220
 cgtagtcata ttccctgaag caaaaaccagg tgccggcggt acaaacagat ctgcgtctcc 35280
 ggtctcgccg cttagatcgc tctgttagt agttgttagta tatccactct ctcaaagcat 35340
 ccaggcgcccc cctggcttcg ggttctatgt aaactccttc atgegcccgt gcccgtataa 35400
 catccaccac cgcaaaaaaa gccacaccca gccaacactac acattcggtc tgcgagtcc 35460
 acacgggagg agcggaaaga gctggaaagaa ccatgtttt ttttttattt caaaagatta 35520
 tccaaaacct caaaaatgaag atcttattaaag tgaacgcgt cccctccgtt ggcgtggtca 35580
 aactctacag ccaaaaaca aataatggca tttgtaaatgttgcataat ggcttccaaa 35640
 aggcaaacgg ccctcacgtc caagtggacg taaaggctaa acccttcagg gtgaatctcc 35700
 tctataaaaca ttccagcacc ttcaaccatg cccaaataat tctatctcg ccaccttctc 35760
 aatatatctc taagcaaaatc cggaaatatta agtccggcca ttgtaaaaat ctgctccaga 35820
 gcgccctcca ccttcagcgtt caagcggcgtt atcatgattt caaaaattca ggttccctcac 35880
 agacctgtat aagattcaaa agcggaaacat taacaaaaat accgcgtatcc cgtaggtccc 35940
 ttccgcaggc cagctgaaca taatcgtgc ggtctgcacg gaccagcgcg gccacttccc 36000
 cggccaggaaacatgacaaaaaa gaacccacac tgattatgac acgcataactc ggagctatgc 36060
 taaccagcgtt agccccgtt taagcttggt gcatggcgg cgatataaaa tgcaagggtgc 36120
 tgctcaaaaaatcaggcaaa gcctcgccgc aaaaagaaaag cacatcgtag tcatgctcat 36180
 gcagataaaag gcaggtaaagc tccggacca ccacagaaaa agacaccatt tttctctcaa 36240
 acatgtctgc gggtttctgc ataaacacaa aataaaataa caaaaaaaaa tttaaacatt 36300
 agaaggctgtt cttacaacag gaaaaaacaac ctttataagc ataagacggc ctacggccat 36360
 gcccggcgttgc cctgtaaaaactggtcacc gtgattaaaa agcaccaccc acagctctc 36420
 ggtcatgtcc ggagtctataa tgtaagactc ggtaaacaca tcaggttgat tcacatcggt 36480
 cagtgttaaa aagcgaccga aatagccccgg gggaaatacat acccgccaggc gtagagacaa 36540
 cattacagcc cccataggag gtataacaaa attaataatggaa gagaaaaaaca cataaacacc 36600
 tgaaaaaaccctcctgcctg gcaaaatagc accctccgc tccagaacaa catacagcgc 36660
 ttccacagcg gcaggccataa cagtcggcct taccgtaaa aagaaaaacc tattaaaaaa 36720
 acaccactcg acacggcacc agctcaatca gtcacagtgt aaaaaaggc caagtgcaga 36780
 gcgagtttat ataggactaa aaaatgacgt aacggttaaa gtccacaaaa aacacccaga 36840
 aaacccgcacg cgaacctacg cccagaaaacg aaagccaaaa aacccacaac ttccctcaat 36900
 cgtcaacttcc gttttccac gttacgtcac ttccatattt aagaaaaacta caattccaa 36960
 cacatacaag ttactccggc ctaaaaccta cgtcaccgc cccgttccca cgccccgcgc 37020
 cacgtcacaactccacccctcattatca tattggcttc aatccaaaat aaggtatatt 37080
 attgtatgtt 37090

<210> 5
 <211> 5955
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> NS cDNA sequence

<221> CDS
 <222> (1)...(5955)

<400> 5		48
atg gcg ccc atc acg gcc tac tcc caa cag acg cgg ggc cta ctt ggt		
Met Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly		
1 5 10 15		
tgc atc atc act agc ctt aca ggc cgg gac aag aac cag gtc gag gga		96
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly		
20 25 30		
gag gtt cag gtg gtt tcc acc gca aca caa tcc ttc ctg gcg acc tgc		144

Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys			
35	40	45	
gtc aac ggc gtg tgt tgg acc gtt tac cat ggt gct ggc tca aag acc		192	
Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr			
50	55	60	
tta gcc ggc cca aag ggg cca atc acc cag atg tac act aat gtg gac		240	
Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp			
65	70	75	80
cag gac ctc gtc ggc tgg cag gcg ccc ccc ggg gcg cgt tcc ttg aca		288	
Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr			
85	90	95	
cca tgc acc tgt ggc agc tca gac ctt tac ttg gtc acg aga cat gct		336	
Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala			
100	105	110	
gac gtc att ccg gtg cgc cg ^g gg ^g ggc gac agt agg ggg agc ctg ctc		384	
Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu			
115	120	125	
tcc ccc agg cct gtc tcc tac ttg aag ggc tct tcg ggt ggt cca ctg		432	
Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu			
130	135	140	
ctc tgc cct tcg ggg cac gct gtg ggc atc ttc cgg gct gcc gta tgc		480	
Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys			
145	150	155	160
acc ccg ggg gtt gcg aag gcg gtg gac ttt gtg ccc gta gag tcc atg		528	
Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met			
165	170	175	
gaa act act atg ccg tct ccg gtc ttc acg gac aac tca tcc ccc ccg		576	
Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro			
180	185	190	
gcc gta ccg cag tca ttt caa gtg gcc cac cta cac gct ccc act ggc		624	
Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly			
195	200	205	
agc ggc aag agt act aaa gtg ccg gct gca tat gca gcc caa ggg tac		672	
Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr			
210	215	220	
aag gtg ctc gtc ctc aat ccg tcc gtt gcc gct acc tta ggg ttt ggg		720	
Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly			
225	230	235	240
g ^c g ^g tat atg tct aag gca cac ggt att gac ccc aac atc aga act ggg		768	
Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly			
245	250	255	

gta agg acc att acc aca ggc gcc ccc gtc aca tac tct acc tat ggc Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly 260 265 270	816
aag ttt ctt gcc gat ggt ggt tgc tct ggg ggc gct tat gac atc ata Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile 275 280 285	864
ata tgt gat gag tgc cat tca act gac tcg act aca atc ttg ggc atc Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile 290 295 300	912
ggc aca gtc ctg gac caa gcg gag acg gct gga gcg cggtt gtc gtg Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val 305 310 315 320	960
ctc gcc acc gct acg cct ccg gga tcg gtc acc gtg cca cac cca aac Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn 325 330 335	1008
atc gag gag gtg gcc ctg tct aat act gga gag atc ccc ttc tat ggc Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly 340 345 350	1056
aaa gcc atc ccc att gaa gcc atc agg ggg gga agg cat ctc att ttc Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe 355 360 365	1104
tgt cat tcc aag aag aag tgc gac gag ctc gcc gca aag ctg tca ggc Cys His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly 370 375 380	1152
ctc gga atc aac gct gtg gcg tat tac cgg ggg ctc gat gtg tcc gtc Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val 385 390 395 400	1200
ata cca act atc gga gac gtc gtt gtc gtg gca aca gac gct ctg atg Ile Pro Thr Ile Gly Asp Val Val Val Ala Thr Asp Ala Leu Met 405 410 415	1248
acg ggc tat acg ggc gac ttt gac tca gtg atc gac tgt aac aca tgt Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys 420 425 430	1296
gtc acc cag aca gtc gac ttc agc ttg gat ccc acc ttc acc att gag Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu 435 440 445	1344
acg acg acc gtg cct caa gac gca gtg tcg cgc tcg cag cgg cgg ggt Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly 450 455 460	1392
agg act ggc agg ggt agg aga ggc atc tac agg ttt gtg act ccg gga Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly 465 470 475 480	1440

gaa cgg ccc tcg ggc atg ttc gat tcc tcg gtc ctg tgt gag tgc tat Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr 485 490 495	1488
gac gcg ggc tgt gct tgg tac gag ctc acc ccc gcc gag acc tcg gtt Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val 500 505 510	1536
agg ttg cgg gcc tac ctg aac aca cca ggg ttg ccc gtt tgc cag gac Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp 515 520 525	1584
cac ctg gag ttc tgg gag agt gtc ttc aca ggc ctc acc cac ata gat His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp 530 535 540	1632
gca cac ttc ttg tcc cag acc aag cag gca gga gac aac ttc ccc tac Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr 545 550 555 560	1680
ctg gta gca taccaa gcc acg gtg tgc gcc agg gct cag gcc cca cct Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro 565 570 575	1728
cca tca tgg gat caaatg tgg aag tgt ctc ata cgg ctg aaa cct acg Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr 580 585 590	1776
ctg cac ggg cca aca ccc ttg ctg tac agg ctg gga gcc gtc caa aat Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn 595 600 605	1824
gag gtc acc ctc acc cac ccc ata acc aaa tac atc atg gca tgc atg Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met 610 615 620	1872
tcg gct gac ctg gag gtc act agc acc tgg gtg ctg gtg ggc gga Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly 625 630 635 640	1920
gtc ctt gca gct ctg gcc gcg tat tgc ctg aca aca ggc agt gtg gtc Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val 645 650 655	1968
att gtg ggt agg att atc ttg tcc ggg agg ccg gct att gtt ccc gac Ile Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp 660 665 670	2016
agg gag ttt ctc tac cag gag ttc gat gaa atg gaa gag tgc gcc tcg Arg Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser 675 680 685	2064
cac ctc cct tac atc gag cag gga atg cag ctc gcc gag caa ttc aag	2112

His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys			
690	695	700	
cag aaa gcg ctc ggg tta ctg caa aca gcc acc aaa caa gcg gag gct			2160
Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala			
705	710	715	720
gct gct ccc gtg gtg gag tcc aag tgg cga gcc ctt gag aca ttc tgg			2208
Ala Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp			
725	730	735	
gcg aag cac atg tgg aat ttc atc agc ggg ata cag tac tta gca ggc			2256
Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly			
740	745	750	
tta tcc act ctg cct ggg aac ccc gca ata gca tca ttg atg gca ttc			2304
Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe			
755	760	765	
aca gcc tct atc acc agc ccg ctc acc acc caa agt acc ctc ctg ttt			2352
Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe			
770	775	780	
aac atc ttg ggg ggg tgg gtg gct gcc caa ctc gcc ccc ccc agc gcc			2400
Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala			
785	790	795	800
gct tcg gct ttc gtg ggc gcc ggc atc gcc ggt gcg gct gtt ggc agc			2448
Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser			
805	810	815	
ata ggc ctt ggg aag gtg ctt gtg gac att ctg gcg ggt tat gga gca			2496
Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala			
820	825	830	
gga gtg gcc ggc gcg ctc gtg gcc ttc aag gtc atg agc ggc gag atg			2544
Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met			
835	840	845	
ccc tcc acc gag gac ctg gtc aat cta ctt cct gcc atc ctc tct cct			2592
Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro			
850	855	860	
ggc gcc ctg gtc ggg gtc gtg tgt gca gca ata ctg cgt cga cac			2640
Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His			
865	870	875	880
gtg ggt ccg gga gag ggg gct gtg cag tgg atg aac cgg ctg ata gcg			2688
Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala			
885	890	895	
ttc gcc tcg cgg ggt aat cat gtt tcc ccc acg cac tat gtg cct gag			2736
Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu			
900	905	910	

agc gac gcc gca gcg cgt gtt act cag atc ctc tcc agc ctt acc atc Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr Ile 915 920 925	2784
act cag ctg ctg aaa agg ctc cac cag tgg att aat gaa gac tgc tcc Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser 930 935 940	2832
aca ccg tgt tcc ggc tcg tgg cta agg gat gtt tgg gac tgg ata tgc Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys 945 950 955 960	2880
acg gtg ttg act gac ttc aag acc tgg ctc cag tcc aag ctc ctg ccg Thr Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro 965 970 975	2928
cag cta ccg gga gtc cct ttt ttc tcg tgc caa cgc ggg tac aag gga Gln Leu Pro Gly Val Pro Phe Ser Cys Gln Arg Gly Tyr Lys Gly 980 985 990	2976
gtc tgg cgg gga gac ggc atc atg caa acc acc tgc cca tgt gga gca Val Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly Ala 995 1000 1005	3024
cag atc acc gga cat gtc aaa aac ggt tcc atg agg atc gtc ggg cct Gln Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro 1010 1015 1020	3072
aag acc tgc agc aac acg tgg cat gga aca ttc ccc atc aac gca tac Lys Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr 1025 1030 1035 1040	3120
acc acg ggc ccc tgc aca ccc tct cca gcg cca aac tat tct agg gcg Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala 1045 1050 1055	3168
ctg tgg cgg gtg gcc gct gag gag tac gtg gag gtc acg cgg gtg ggg Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly 1060 1065 1070	3216
gat ttc cac tac gtg acg ggc atg acc act gac aac gta aag tgc cca Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro 1075 1080 1085	3264
tgc cag gtt ccg gct cct gaa ttc ttc acg gag gtg gac gga gtg cgg Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg 1090 1095 1100	3312
ttg cac agg tac gct ccg gcg tgc agg cct ctc cta cgg gag gag gtt Leu His Arg Tyr Ala Pro Ala Cys Arg Pro Leu Leu Arg Glu Glu Val 1105 1110 1115 1120	3360
aca ttc cag gtc ggg ctc aaccaa tac ctg gtt ggg tca cag cta cca Thr Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro 1125 1130 1135	3408

tgc gag ccc gaa ccg gat gta gca gtg ctc act tcc atg ctc acc gac Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp 1140 1145 1150	3456
ccc tcc cac atc aca gca gaa acg gct aag cgt agg ttg gcc agg ggg Pro Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly 1155 1160 1165	3504
tct ccc ccc tcc ttg gcc agc tct tca gct agc cag ttg tct gcg cct Ser Pro Pro Ser Leu Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro 1170 1175 1180	3552
tcc ttg aag gcg aca tgc act acc cac cat gtc tct ccg gac gct gac Ser Leu Lys Ala Thr Cys Thr His His Val Ser Pro Asp Ala Asp 1185 1190 1195 1200	3600
ctc atc gag gcc aac ctc ctg tgg cgg cag gag atg ggc ggg aac atc Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile 1205 1210 1215	3648
acc cgc gtg gag tcg gag aac aag gtg gta gtc ctg gac tct ttc gac Thr Arg Val Glu Ser Glu Asn Lys Val Val Val Leu Asp Ser Phe Asp 1220 1225 1230	3696
ccg ctt cga gcg gag gag gat gag agg gaa gta tcc gtt ccg gcg gag Pro Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu 1235 1240 1245	3744
atc ctg cgg aaa tcc aag aag ttc ccc gca gcg atg ccc atc tgg gcg Ile Leu Arg Lys Ser Lys Lys Phe Pro Ala Ala Met Pro Ile Trp Ala 1250 1255 1260	3792
ccg ccg gat tac aac cct cca ctg tta gag tcc tgg aag gac ccg gac Arg Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp 1265 1270 1275 1280	3840
tac gtc cct ccg gtg gtg cac ggg tgc ccg ttg cca cct atc aag gcc Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala 1285 1290 1295	3888
cct cca ata cca cct cca ccg aga aag agg acg gtt gtc cta aca gag Pro Pro Ile Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu 1300 1305 1310	3936
tcc tcc gtg tct tcc gcc tta gcg gag ctc gct act aag acc ttc ggc Ser Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly 1315 1320 1325	3984
agc tcc gaa tca tcg gcc gtc gac agc ggc acg gcg acc gcc ctt cct Ser Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro 1330 1335 1340	4032
gac cag gcc tcc gac gac ggt gac aaa gga tcc gac gtt gag tcg tac	4080

Asp Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr			
1345	1350	1355	1360
tcc tcc atg ccc ccc ctt gag ggg gaa ccg ggg gac ccc gat ctc agt			4128
Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser			
1365	1370	1375	
gac ggg tct tgg tct acc gtg agc gag gaa gct agt gag gag gat gtc gtc			4176
Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val			
1380	1385	1390	
tgc tgc tca atg tcc tac aca tgg aca ggc gcc ttg atc acg cca tgc			4224
Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys			
1395	1400	1405	
gct gcg gag gaa agc aag ctg ccc atc aac gcg ttg agc aac tct ttg			4272
Ala Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu			
1410	1415	1420	
ctg cgc cac cat aac atg gtt tat gcc aca aca tct cgc agc gca ggc			4320
Leu Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly			
1425	1430	1435	1440
ctg cgg cag aag gtc acc ttt gac aga ctg caa gtc ctg gac gac			4368
Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp			
1445	1450	1455	
cac tac cgg gac gtg ctc aag gag atg aag gcg aag gcg tcc aca gtt			4416
His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val			
1460	1465	1470	
aag gct aaa ctc cta tcc gta gag gaa gcc tgc aag ctg acg ccc cca			4464
Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro			
1475	1480	1485	
cat tcg gcc aaa tcc aag ttt ggc tat ggg gca aag gac gtc cgg aac			4512
His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn			
1490	1495	1500	
cta tcc agc aag gcc gtt aac cac atc cac tcc gtg tgg aag gac ttg			4560
Leu Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu			
1505	1510	1515	1520
ctg gaa gac act gtg aca cca att gac acc acc atc atg gca aaa aat			4608
Leu Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn			
1525	1530	1535	
gag gtt ttc tgt gtc caa cca gag aaa gga ggc cgt aag cca gcc cgc			4656
Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg			
1540	1545	1550	
ctt atc gta ttc cca gat ctg gga gtc cgt gta tgc gag aag atg gcc			4704
Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala			
1555	1560	1565	

ctc tat gat gtg gtc tcc acc ctt cct cag gtc gtg atg ggc tcc tca Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser 1570 1575 1580	4752
tac gga ttc cag tac tct cct ggg cag cga gtc gag ttc ctg gtg aat Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn 1585 1590 1595 1600	4800
acc tgg aaa tca aag aaa aac ccc atg ggc ttt tca tat gac act cgc Thr Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg 1605 1610 1615	4848
tgt ttc gac tca acg gtc acc gag aac gac atc cgt gtt gag gag tca Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser 1620 1625 1630	4896
att tac caa tgt tgt gac ttg gcc ccc gaa gcc aga cag gcc ata aaa Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys 1635 1640 1645	4944
tcg ctc aca gag cgg ctt tat atc ggg ggt cct ctg act aat tca aaa Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys 1650 1655 1660	4992
ggg cag aac tgc ggt tat cgc cgg tgc cgc gcg agc ggc gtg ctg acg Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr 1665 1670 1675 1680	5040
act agc tgc ggt aac acc ctc aca tgt tac ttg aag gcc tct gca gcc Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala 1685 1690 1695	5088
tgt cga gct gcg aag ctc cag gac tgc acg atg ctc gtg aac gga gac Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp 1700 1705 1710	5136
gac ctt gtc gtt atc tgt gaa agc gcg gga acc caa gag gac gcg gcg Asp Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala 1715 1720 1725	5184
agc cta cga gtc ttc acg gag gct atg act agg tac tct gcc ccc ccc Ser Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro 1730 1735 1740	5232
ggg gac ccg ccc caa cca gaa tac gac ttg gag ctg ata aca tca tgt Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys 1745 1750 1755 1760	5280
tcc tcc aat gtg tcg gtc gcc cac gat gca tca ggc aaa agg gtg tac Ser Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr 1765 1770 1775	5328
tac ctc acc cgt gat ccc acc acc ccc ctc gca cgg gct gcg tgg gaa Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu 1780 1785 1790	5376

aca gct aga cac act cca gtt aac tcc tgg cta ggc aac att atc atg Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met 1795 1800 1805	5424
tat gcg ccc act ttg tgg gca agg atg att ctg atg act cac ttc ttc Tyr Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe 1810 1815 1820	5472
tcc atc ctt cta gca cag gag caa ctt gaa aaa gcc ctg gac tgc cag Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln 1825 1830 1835 1840	5520
atc tac ggg gcc tgt tac tcc att gag cca ctt gac cta cct cag atc Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile 1845 1850 1855	5568
att gaa cga ctc cat ggc ctt agc gca ttt tca ctc cat agt tac tct Ile Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser 1860 1865 1870	5616
cca ggt gag atc aat agg gtg gct tca tgc ctc agg aaa ctt ggg gta Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val 1875 1880 1885	5664
cca ccc ttg cga gtc tgg aga cat cgg gcc agg agc gtc cgcc gct agg Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg 1890 1895 1900	5712
cta ctg tcc cag ggg agg gcc act tgt ggc aag tac ctc ttc Leu Leu Ser Gln Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe 1905 1910 1915 1920	5760
aac tgg gca gtg aag acc aaa ctc aaa ctc act cca atc ccg gct gcg Asn Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala 1925 1930 1935	5808
tcc cag ctg gac ttg tcc ggc tgg ttc gtt gct tac agc ggg gga Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly 1940 1945 1950	5856
gac ata tat cac agc ctg tct cgt gcc cga ccc cgcc tgg ttc atg ctg Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu 1955 1960 1965	5904
tgc cta ctc cta ctt tct gta ggg gta ggc atc tac ctg ctc ccc aac Cys Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn 1970 1975 1980	5952
cga Arg 1985	5955
<210> 6	
<211> 1984	

<212> PRT

<213> Artificial Sequence

<220>

<223> NS sequence

<400> 6
Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys
1 5 10 15
Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu
20 25 30
Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val
35 40 45
Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu
50 55 60
Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln
65 70 75 80
Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro
85 90 95
Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp
100 105 110
Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser
115 120 125
Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu
130 135 140
Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr
145 150 155 160
Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu
165 170 175
Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala
180 185 190
Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser
195 200 205
Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys
210 215 220
Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala
225 230 235 240
Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val
245 250 255
Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys
260 265 270
Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile
275 280 285
Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly
290 295 300
Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu
305 310 315 320
Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile
325 330 335
Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys
340 345 350
Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys
355 360 365
His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu
370 375 380

Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile
 385 390 395 400
 Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
 405 410 415
 Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val
 420 425 430
 Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr
 435 440 445
 Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg
 450 455 460
 Thr Gly Arg Gly Arg Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu
 465 470 475 480
 Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp
 485 490 495
 Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg
 500 505 510
 Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His
 515 520 525
 Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala
 530 535 540
 His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu
 545 550 555 560
 Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro
 565 570 575
 Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu
 580 585 590
 His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu
 595 600 605
 Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser
 610 615 620
 Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val
 625 630 635 640
 Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val Ile
 645 650 655
 Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp Arg
 660 665 670
 Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser His
 675 680 685
 Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys Gln
 690 695 700
 Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala Ala
 705 710 715 720
 Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp Ala
 725 730 735
 Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu
 740 745 750
 Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr
 755 760 765
 Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe Asn
 770 775 780
 Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala Ala
 785 790 795 800
 Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser Ile
 805 810 815
 Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly

	820	825	830
Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met Pro			
835	840	845	
Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly			
850	855	860	
Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val			
865	870	875	880
Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe			
885	890	895	
Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser			
900	905	910	
Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr Ile Thr			
915	920	925	
Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser Thr			
930	935	940	
Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys Thr			
945	950	955	960
Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro Gln			
965	970	975	
Leu Pro Gly Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys Gly Val			
980	985	990	
Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly Ala Gln			
995	1000	1005	
Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro Lys			
1010	1015	1020	
Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr Thr			
1025	1030	1035	1040
Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala Leu			
1045	1050	1055	
Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly Asp			
1060	1065	1070	
Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro Cys			
1075	1080	1085	
Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg Leu			
1090	1095	1100	
His Arg Tyr Ala Pro Ala Cys Arg Pro Leu Leu Arg Glu Glu Val Thr			
1105	1110	1115	1120
Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro Cys			
1125	1130	1135	
Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro			
1140	1145	1150	
Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly Ser			
1155	1160	1165	
Pro Pro Ser Leu Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser			
1170	1175	1180	
Leu Lys Ala Thr Cys Thr Thr His His Val Ser Pro Asp Ala Asp Leu			
1185	1190	1195	1200
Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr			
1205	1210	1215	
Arg Val Glu Ser Glu Asn Lys Val Val Val Leu Asp Ser Phe Asp Pro			
1220	1225	1230	
Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu Ile			
1235	1240	1245	
Leu Arg Lys Ser Lys Lys Phe Pro Ala Ala Met Pro Ile Trp Ala Arg			
1250	1255	1260	

Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp Tyr
 1265 1270 1275 1280
 Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala Pro
 1285 1290 1295
 Pro Ile Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu Ser
 1300 1305 1310
 Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly Ser
 1315 1320 1325
 Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro Asp
 1330 1335 1340
 Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr Ser
 1345 1350 1355 1360
 Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp
 1365 1370 1375
 Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val Cys
 1380 1385 1390
 Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala
 1395 1400 1405
 Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu
 1410 1415 1420
 Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly Leu
 1425 1430 1435 1440
 Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp His
 1445 1450 1455
 Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys
 1460 1465 1470
 Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro His
 1475 1480 1485
 Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu
 1490 1495 1500
 Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu Leu
 1505 1510 1515 1520
 Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu
 1525 1530 1535
 Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu
 1540 1545 1550
 Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu
 1555 1560 1565
 Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser Tyr
 1570 1575 1580
 Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Thr
 1585 1590 1595 1600
 Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys
 1605 1610 1615
 Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser Ile
 1620 1625 1630
 Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys Ser
 1635 1640 1645
 Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly
 1650 1655 1660
 Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr
 1665 1670 1675 1680
 Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala Cys
 1685 1690 1695

Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp Asp
 1700 1705 1710
 Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala Ser
 1715 1720 1725
 Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly
 1730 1735 1740
 Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser
 1745 1750 1755 1760
 Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr Tyr
 1765 1770 1775
 Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr
 1780 1785 1790
 Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr
 1795 1800 1805
 Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser
 1810 1815 1820
 Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln Ile
 1825 1830 1835 1840
 Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile
 1845 1850 1855
 Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro
 1860 1865 1870
 Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val Pro
 1875 1880 1885
 Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu
 1890 1895 1900
 Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn
 1905 1910 1915 1920
 Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser
 1925 1930 1935
 Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly Asp
 1940 1945 1950
 Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu Cys
 1955 1960 1965
 Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1970 1975 1980

<210> 7
 <211> 4909
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pV1J nucleic acid

<400> 7
 tcgcgcgtt cggtgatgac ggtaaaaacc tctgacacat gcagctcccg gagacggtca 60
 cagcttgct gtaaggcgat gcccggagca gacaagcccg tcagggcgcg tcagcgggtg 120
 ttggcgggtg tcggggctgg cttaactatg cggcataga gcagattgta ctgagagtgc 180
 accatatgcg gtgtgaaata ccgcacagat gcgtaaaggag aaaataccgc atcagatgg 240
 ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg 300
 tccaacatta ccgcacatgtt gacattgatt attgactagt tattaatagt aatcaattac 360
 ggggtcattt gttcatagcc catatatggg gttccgcgtt acataactta cggtaaatgg 420
 cccgcctggc tgaccggcca acgacccccc cccattgacg tcaataatga cgtatgtcc 480
 catagtaacg ccaatagggc ctttcattt acgtcaatgg gtggagtatt tacggtaaac 540

<210> 8
<211> 35935
<212> DNA
<213> Adenovirus serotype 6

tttctgtggg gctcatccca ggcaaagtta gtctgcagaa ttaaggagga ttacaagtgg	1860
gaatttgaag agctttgaa atcctgtggt gagctgtttt attcttgaa tctgggtcac	1920
caggcgctt tccaagagaa ggtcatcaag actttggatt ttccacacc ggggcgcgct	1980
gcggctgtcg ttgcctttt gagtttata aaggataaat ggagcgaaga aacccatctg	2040
agcgggggggt acctgctgga tttctggcc atgcacatgt ggagagcggt tgtgagacac	2100
aagaatcgcc tgctactgtt gtcttcgcgtaataaccgac ggaggagcag	2160
cagcagcagc aggaggaagc caggcggcgg cgccaggagc agagcccatg gaacctgaga	2220
gccggcctgg accctcgga atgaatgtt tacaggtggc tgaactgtat ccagaactga	2280
gacgcattt gacaattaca gaggatgggc agggctaaa ggggtaaag agggagcggg	2340
gggcttgtga ggctacagag gaggcttagga atctagctt tagcttaatg accagacacc	2400
gtcctgagtg tattactttt caacagatca aggataattt cgctaatttgcgatctgc	2460
tggcgcagaa gtattccata gaggcgttgc ccacttacttgcgatccca ggggatgatt	2520
ttgaggaggc tattaggta tatgcaaagg tggacttag gccagattgc aagtacaaga	2580
tcagcaaact tgtaaatatc aggaatttgtt gctacatttgcgatcc tggaaacggg gcccgggtgg	2640
agatagatc ggaggatagg gtggccttta gatgttagcat gataaatatg tggccggggg	2700
tgcttggcat ggacgggtt gttattatga atgttaagggt tactggcccc aatttttagcg	2760
gtacggttt cctggccat accaaccttta tcctacacgg tctaagcttgcgatcc tattgggttta	2820
acaatacctg tggaaagcc tggaccgatg taagggttcg gggctgtgccc ttactgtct	2880
gcttggaaagggg ggtgggtgtt cggccaaaaa gcaggcttc aattaagaaa tgccttttgc	2940
aaaggtgtac ctgggttgc ctgtctgagg gtaactccag ggtgcgcac aatgtggcct	3000
ccgactgtgg ttgcttcatg ctatgtaaaaa gctgtgtgtt gattaagcat aacatgttat	3060
gtggcaactg cgaggacagg gcctctcaga tgctgacccgcgatcc tccggacggc aactgttacc	3120
tgctgaagac cattcacgttgc ccagccact ctcgcaaggc ctggccagtg tttgagcata	3180
acatactgac ccgctgttcc ttgcatttgg gtaacaggag ggggggttcc ctacccattacc	3240
aatgcaattt gagtacact aagatatttgc ttgaggccga gacatgtcc aaggtgaacc	3300
tgaacgggggt gtttgcacatg accatgaaga tcttggaaagggt gctgaggta gatgagaccc	3360
gcaccagggtt cagaccctgc gagggtggcg gtaaacatata taggaaccag cctgtgtatgc	3420
tggatgttgcgatcc cgaggagcttgcgatcc acttgggtgtt ggcctgcacc cgcgtctgagt	3480
ttggctcttagt cgatgaagat acagatttgc gtaactgaaat gtgtggcggtt ggcttaaggg	3540
tggggaaagaa tatataaggt gggggcttta ttttttttatcttgcgatcc ttttttttttgc	3600
ccggcccat gaggccaaac tcgtttgtatg gaaaggatgtt gacatgttgcgatcc tttgacaacgc	3660
gcacggccat gggccgggggttgcgatcc ttttttttttgcgatcc ttttgcgatcc ttttgcgatcc	3720
ccgtctctgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	3780
agactgcaggc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	3840
actttgcgtttt ctttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	3900
acaagtttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	3960
ctcagcagtttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4020
atgcgggttta aaacataaat aaaaaaccatgacttgcgatcc ttttgcgatcc ttttgcgatcc	4080
cttgctgtcttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4140
cggttgggttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4200
acatggggcat aaggccgttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4260
gggttgggttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4320
ttttcagtagt caagctgttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4380
taagctggatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4440
tggctatgttcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4500
tgtatccgttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4560
tggagacggcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4620
gcccacggcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4680
ccaggatgttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4740
gtataatgttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4800
ctttgagttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4860
gggttggggatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4920
cggttggggcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	4980
tgccgtcatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	5040
ccctgaccatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc ttttgcgatcc	5100

caaagggttt	caacgggttg	agaccgtccg	ccgttaggcatt	gttttgcgc	gtttgacc	5160
gcagttccag	gcggccccac	agctcggtca	cctgcctcac	ggcatctcg	tccagcatat	5220
ctcctcggtt	cgcgggttgg	ggcggcttc	gctgtacggc	agtagtcgg	gctcgccag	5280
acggggccagg	gtcatgttctt	tccacggggc	cagggcctc	gtcagcgtag	tctgggtcac	5340
ggtaaggggg	tgcgtcccg	gctgcgcgt	ggccagggtg	cgcttgggg	tggccctgct	5400
ggtgcgtaaag	cgctgccgt	cttcgcctc	cgcgccggc	aggtgcatt	tgaccatgg	5460
gtcatagtcc	agcccctccg	cggcgtggc	cttggcgcc	agcttgcct	tggaggaggc	5520
gcccacccag	gggcagtgc	gacttttag	ggcgtagagc	ttgggcgcga	gaaataccga	5580
ttccggggag	taggcattccg	cgccgcaggc	cccgacagac	gtctcgcat	ccacgagcca	5640
ggtgagctct	ggccgttcgg	ggtaaaaaac	cagggttccc	ccatgcctt	tgatgcgtt	5700
cttacctctg	gtttccatga	gccgggttcc	acgctcggt	acgaaaaggc	tgtccgtgtc	5760
cccgtataca	gacttgagag	gcctgcctc	gaggggtt	ccgcggctct	cctcgatag	5820
aaactcgac	cactctgaga	caaaggctcg	cgtccaggcc	agcacgaagg	aggctaagt	5880
ggaggggtag	cggtcgttgt	ccactagggg	gtccactcgc	tccagggtgt	gaagacacat	5940
gtcgccctct	tcggcatcaa	ggaaggtat	tggttttag	gttgcggcca	cgtgaccggg	6000
tgttccctgaa	ggggggctat	aaaaggggtt	gggggcgcgt	tcgtccctac	tctctccgc	6060
atcgctgtct	gcgagggcca	gctgttggg	tgagtactcc	ctctgaaaag	cgggcatgac	6120
ttctgcgtta	agattgtcag	tttccaaaaa	cgaggaggat	ttgatattca	cctggcccgc	6180
ggtgatgcct	ttgaggggtt	ccgcattccat	ctggtcagaa	aagacaatct	ttttgtgtc	6240
aagcttggtg	gcaaacgacc	cgtagagggc	gttggacagc	aacttggcga	tggagcgcag	6300
ggtttgggtt	ttgtcgcat	cggcgcgc	cttggccgcg	atgttttagct	gcacgtattc	6360
gcgcgcac	caccgcatt	cgggaaagac	ggtggcgc	tcgtccggca	ccaggtgcac	6420
gcgcacccg	cgggtgtca	gggtgacaag	gtcaacgc	gtggctacct	ctccgcgtag	6480
gcgcgttgc	gtccagcaga	ggcggccg	cttgcgcag	cagaatggcg	gtaggggtc	6540
tagctcg	tcgtccgggg	ggtctcg	cacggtaaaag	accccccggca	gcaggcgc	6600
gtcgaagtag	tctatcttgc	atccttgc	gtctagcgc	tgctccat	cggggccgc	6660
aagcgcgcgc	tcgtatgggt	tgagtgggg	accccatggc	atgggggtgg	tgagcgcgga	6720
ggcgtacatg	ccgcacatgt	cgtaaacgta	gaggggtct	ctgagttattc	caagatgt	6780
agggttagcat	cttccaccgc	ggatgctgg	gcmcacgta	ttgtatagtt	cgtgcgaggg	6840
agcgaggagg	tcgggaccga	ggttgctac	ggcgggctgc	tctgcgcg	agactatct	6900
cctgaagatg	gcatgtgat	tggatgat	ggttggacgc	tgaagacgt	tgaagtcgc	6960
gtctgtgaga	cctaccgcgt	cacgcac	ggaggcgtag	gagtcgcgc	gcttggtgc	7020
cagctcg	gtgacccgtca	cgtttaggg	gcaagtgc	agggtttcot	tgtatgtgc	7080
atacttatcc	tgtccctt	ttttccacag	ctcgcgg	aggacaaact	cttcgcgg	7140
tttccagtag	tcttggatcg	gaaacccgc	ggcctccgaa	cgtaagagc	ctagcatgt	7200
gaactgggt	acggcctgtt	aggcgcagca	tccctttct	acgggttagc	cgtatgcct	7260
cgcggcc	eggagcgg	tgtgggtgag	cgcaaaagg	tccctgacca	tgacttgc	7320
gtactggat	ttgaagtcag	tgtcgtcg	tccccc	tcccagagc	aaaagtccgt	7380
gcgcgtttt	gaacgcggat	ttggcagggc	gaagggtaca	tcgttgaaga	gtatcttcc	7440
cgcgcgaggc	ataaaagttc	gtgtgatgc	gaagggtccc	ggcacctcg	aacgggttt	7500
aattacctgg	gcccggagca	cgatctcg	aaaggcgtt	atgttgcgc	ccacaatgt	7560
aagttccaa	aagcgcggga	tgccttgc	ggaaggcaat	tttttaagtt	cctcgtaggt	7620
gagctctca	ggggagctga	gcccgtc	tgaaaggccc	cagtctgc	gatgagggtt	7680
ggaagcgcac	aatagagctcc	acaggtc	ggcattagc	atttgcagg	gtcgcgaaa	7740
ggtccta	ttggcgcac	tggccattt	ttctgggt	atgcagtaga	agtaagcgg	7800
gtcttgc	cagcggtccc	atccaaagg	cgccgtc	tctcg	cagtca	7860
aggctcatct	ccgcgcact	tcatgacc	catgaaggc	acgagct	tcccaaggc	7920
ccccatccaa	gtataggtct	ctacatcg	ggtgcaca	agacgc	tgcgaggat	7980
cgagccgatc	gggaaagaact	ggatctcc	ccaccaatt	gaggagtgg	tattgtgt	8040
gtgaaagtag	aagtccctgc	gacggccga	acactcg	ttggctttt	aaaaacgtgc	8100
gcagta	cagcggtgc	cggtctgtac	atcctgcac	aggttgcac	gacgaccgc	8160
cacaaggaag	cagagtggga	atttgcagcc	ctcgcctgg	gggttgggt	ggtgttctc	8220
tacttcggct	gcttgcctt	gaccgtctgg	ctgctcg	ggagttacgg	tggatcggac	8280
caccacgccc	cgcgcggccca	aagtccagat	gtccgcgc	ggcggtcgg	gcttgcgtac	8340
aacatcgcc	agatgggagc	tgtccat	ctggagctc	cgcggcg	ggtcaggcgg	8400

gagctcctgc	aggtttacct	cgcata	tagacg	ggtc	cagg	gcg	cgg	ctagat	ccagg	tata	gtata	8460
cctaatttcc	aggggcttgt	ttgtggcg	gtcgat	ggct	tgca	aa	agg	gc	catcccc	8520		
ccggcgc	act	acggta	ccgc	gcgg	cgg	gtgg	ccgc	gggt	gtcc	8580		
atctaaa	agc	gtgac	cg	cg	gag	cccc	ggag	gtt	ccgg	8640		
agaggggg	ca	gggcac	gtc	ggc	ccc	gc	ggg	cagg	gtct	8700		
ttgctgg	ga	acgc	acgac	ggg	cgg	gtt	g	ctgg	cgct	8760		
acgacgg	cc	cggt	gag	ctt	gac	tttt	gat	caat	ttcg	8820		
ttgacgg	cc	cctgg	gca	aa	atct	ccct	tc	tttgc	tgc	8880		
tcggccat	ga	act	gtc	cat	ctt	cc	tgg	atct	ccac	8940		
gtggcgg	ga	gtt	gtt	gga	aat	gcgg	at	gag	gtcg	9000		
tcgttcc	aga	cg	cg	ctgt	ta	gacc	ac	ggccc	cc	9060		
tgcg	cg	ag	at	ccac	ttt	ccgg	cat	cg	ggcg	9120		
aggtag	ttt	g	gggt	gttgc	gtt	gtt	cc	aca	at	9180		
aacgtgg	att	cg	ttt	at	atc	cccc	aa	agg	gtcg	9240		
acggc	aa	gt	aaaa	act	g	ggac	ac	gg	cc	9300		
cgat	gag	ct	ggc	ac	gt	ccgc	acc	ttt	ctt	9360		
tctt	ttt	ca	tctt	ttt	ca	ata	gg	ttt	ctt	9420		
ggaggggg	ga	cac	ggc	g	ac	ggc	gg	gg	gg	9480		
atct	cccc	gc	gac	gg	ttt	ccgg	gc	gtt	ccgc	9540		
agtt	ggaa	ga	ccc	cc	ttt	gggtt	gg	gggg	gg	9600		
agggata	acgg	cg	cta	ac	ttt	gggtt	gg	gggg	gg	9660		
gac	ctg	ag	cc	cg	atc	ggggat	cc	gg	gg	9720		
tcac	ag	tc	gg	gtt	gtt	gggt	cc	gg	gg	9780		
ttt	ctgg	gg	gg	gtt	gtt	gtt	cc	gg	gg	9840		
gtcg	ac	aga	aa	gtt	gtt	gtt	cc	gg	gg	9900		
cccc	agg	ttt	gaca	tc	gg	gggt	cc	gg	gg	9960		
accgg	ca	ttt	ctt	ttt	gtt	ccgt	cc	gtt	cc	10020		
gcgg	gg	ttt	cc	cc	gtt	ccgt	cc	gtt	cc	10080		
ctc	at	cg	gt	cc	ttt	cc	cc	gtt	cc	10140		
ac	ct	cg	gt	cc	ttt	cc	cc	gtt	cc	10200		
ttgat	gggt	ttt	ttt	ttt	ttt	ttt	ttt	ttt	ttt	10260		
gag	ag	tc	cc	gg	ttt	ttt	ttt	ttt	ttt	10320		
gtcc	gc	ac	cc	ca	ttt	ttt	ttt	ttt	ttt	10380		
cagc	gt	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	10440		
tagat	gtt	cc	cc	cc	ttt	ttt	ttt	ttt	ttt	10500		
cg	gg	cc	cc	cc	ttt	ttt	ttt	ttt	ttt	10560		
ccgg	tc	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	10620		
ggc	act	ct	gtt	cc	ttt	ttt	ttt	ttt	ttt	10680		
tcg	ag	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	10740		
cagg	gt	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	10800		
gct	gt	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	10860		
gcg	aa	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	10920		
ggc	gg	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	10980		
ccgt	cc	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	11040		
ttt	cc	cc	cc	cc	ttt	ttt	ttt	ttt	ttt	11100		
ca	ag	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	11160		
ac	at	cc	cc	cc	ttt	ttt	ttt	ttt	ttt	11220		
ca	ct	cc	cc	cc	ttt	ttt	ttt	ttt	ttt	11280		
cg	gt	cc	cc	cc	ttt	ttt	ttt	ttt	ttt	11340		
ct	ttt	cc	cc	cc	ttt	ttt	ttt	ttt	ttt	11400		
gg	gg	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	11460		
cc	cc	gg	gg	gg	ttt	ttt	ttt	ttt	ttt	11520		
ac	cc	ca	cc	cc	ttt	ttt	ttt	ttt	ttt	11580		
gt	gt	ac	cc	cc	ttt	ttt	ttt	ttt	ttt	11640		
gt	ac	cc	cc	cc	ttt	ttt	ttt	ttt	ttt	11700		

gtgcagcaca gcagggacaa cgaggcattc agggatgcgc tgctaaacat agtagagccc	11760
gagggccgct ggctgctcgta ttgataaac atcctgcaga gcatacggtt gcaggagcgc	11820
agcttgagcc tggctgacaa ggtggccgc atcaactatt ccatgcttag cctggcaag	11880
tttacccc gcaagatata ccatacccc tacgttcca tagacaagga gttaaagatc	11940
gaggggttct acatgcgcat ggcgctgaag gtgcctacat tgagcgcga cctggcggtt	12000
tatcgcaacg agcgcattca caaggccgtg agcgtgagcc ggcggcgcga gctcagcgac	12060
cgcgagctga tgcacagcct gcaaaggccc ctggctggca cggcagcgg cgatagagag	12120
gccgagtctt actttgacgc gggcgtgac ctgcgttggg ccccaagccg acgcgcctg	12180
gagggcagctg gggccggacc tggctggc gtggcaccccg cgcgcgtgg caacgtcggc	12240
ggcgtggagg aatatgacga ggacgatgag tacgaggccag aggacggcga gtactaaagc	12300
gtgtatgttc tgatcagat atgacagac caacggaccc ggcgggtgcg gcggcgctgc	12360
agagccagcc gtccggcctt aactccacgg acgactggcg ccaggtcatg gaccgcata	12420
tgtcgctgac tgcgcgcaat cctgcgcgt tccggcagca gccgcaggcc aaccggctct	12480
ccgcaattct ggaagcgggtg gtccggcgcg ggcggaaaccc caccgcacgg aagggtgtgg	12540
cgatcgtaaa cgcgcgtggcc gaaaacaggg ccatccggcc cgacgaggcc ggcctggct	12600
acgacgcgct gcttacgcgc gtggctcgat acaacagccg caacgtgcg accaacctgg	12660
accggcttgtt gggggatgtg cgcgaggccg tggcgcagcg tgagcgcgcg cagcagcagg	12720
gcaacacttggg ctccatgtt gcaactaaac ctttcctgag tacacagccc gccaacgtgc	12780
cgcggggaca ggaggactac accaactttg tgagcgcact gcggctaatt gtgactgaga	12840
caccgcggaa tgagggtgtac cagtctgggc cagactattt tttccagacc agtagacaag	12900
gcctgcagac cgtaaacctg agccaggctt tcaaaaactt gcaaggggctg tgggggtgc	12960
gggtccccac aggccgaccgc ggcgaccgtt ctgcgtgt gacgcggcc acgcgcctgt	13020
tgctgtgtct aatagcgcctt ttcacggaca gtggcagcgt gtccgggac acatacctag	13080
gtcacttgct gacactgtac cgcgaggcca taggtcaggc gcatgtggac gaggcatactt	13140
tccaggagat tacaagtgtc agccgcgcgc tggggcagga ggacacgggc agcctggagg	13200
caaccctaaa ctacctgctg accaaccggc ggcagaagat cccctcggt cacagttaa	13260
acagcgagga ggagcgcatt ttgcgttacg tgcagcagag cgtgaggcctt aacctgtatgc	13320
gcgcacggggt aacgcggccgc gtggcgttgc acatgaccgc ggcacacatg gaaccggca	13380
tgtatgcctt aaaccggccg tttatcaacc gcctaattggc ctacttgcac cgcgcggccg	13440
ccgtgaaccc cgagtatttc accaatgcctt tcttgaaccc gcactggcta cccggccctg	13500
gtttctacac cgggggattt cgggttcccg agggttaacga tggattccctc tgggacgaca	13560
tagacgacag cgtgttttcc cgcacccgcg agaccctgtt agatgtgca cagcgcgagc	13620
aggcagagggc ggcgtgcga aaggaaagct tccgcaggcc aagcagcttgc tccgatctag	13680
gcgcgtgcgc cccgcgggtca gatgcgttgc gcccatttcc aagcttgcata gggctcttta	13740
ccagcactcg caccacccgc cgcgcctgc tggcggagga ggagtaccta aacaactcgc	13800
tgctgcagcc gcagcgcgaa aaaaacctgc ctccggcatt tcccaacaac gggatagaga	13860
gcctagtggc caagatgagt agatggaaaga cgtacgcgcg ggcgcacagg gacgtgcac	13920
gcccgcgccc gcccacccgt cgtcaaaaggc acgaccgtca ggggggtctg gtgtggagg	13980
acgatgactc ggcagacgcgac agcagcgttcc tggatttttttggg aggagtggc aacccttttgc	14040
cgcaccttcg ccccgaggctg gggagaatgt tttaaaaaaaa aaaaagcatg atgc当地	14100
aaaaactcac caaggccatg gcaccgcgcg ttgggtttct tttatcccccc ttagtatgcg	14160
gcgcgcggcg atgtatgagg aagggttccctc tccctccatc gagagtgtgg tgagcgcggc	14220
gccagtggcg gcccgcgtt gttctccctt cgatgcctt ctggacccgc cgtttgcc	14280
tccgcgttac ctgcggccta cccggggggag aaacagcatc cgttactctg agttggcacc	14340
cctattcgac accacccgtg tttatcggtt ggacaacaag tcaacggatg tggcatccct	14400
gaactaccatg aacgaccaca gcaacttttgc gaccacggc attcaaaaaca atgactacag	14460
cccgggggag gcaagcacac agaccatcaa tcttgacgcgc cggtcgcact gggcggcga	14520
cctgaaaacc atcctgcata ccaacatgcgc aaatgtgaac gagttcatgt ttaccaataa	14580
gtttaaaggcg cgggtgttgc tgtcgcgtt gcctactaag gacaatcagg tggagctgaa	14640
atacgagttgg gtggagttca cgtgcgcgcg gggcaactac tccgcgcacca tgaccataga	14700
ccttatgaac aacgcgtatcg tggagacta cttgaaatgt ggcagacaga acgggggttct	14760
ggaaagcgac atcggggtaa agtttgacac cgcacacttcc agactgggtt ttgacccgt	14820
cactggtctt gtcacgcctt gggatataac aaacgaagcc ttccatccag acatcattt	14880
gctgccagga tgcgggggtgg acttcacccca cagccgcctg agcaacttgc tggcgcac	14940
caagcggcaa cccttcagg agggttttag gatcacctac gatgtatgttggggttaa	15000

cattcccgca ctgttggatg tggacgccta ccaggcgcagc ttgaaagatg acaccgaaca
gggcgggggt ggcgcaggcg gcagcaacag cagtggcagc ggccggaaag agaactccaa
cgccgcagcc gcggcaatgc agccgtgga ggacatgaac gatcatgccca ttccggcgaa
cacctttgcc acacgggctg aggagaaggcg cgctgaggcc gaagcagcgg ccgaagctgc
cgcccccgct gcgc当地cccg aggtcgagaa gcctcagaag aaaccgggtga tcaaaccct
gacagaggac agcaagaaac gcagttacaa cctaataagc aatgacagca cttcacc
gtaccgcagc tggtaacctg catacaacta cggc当地ccct cagaccggaa tccgctcatg
gaccctgctt tgcaactctg acgtaacctg cggctcggag caggctact ggtcgttgc
agacatgatg caagacccc tgacccctcg ctccacgc cagatcagca actttccgt
ggtggcgcc gagctgttgc ccgtcactc caagagctt tacaacgacc agggccgtcta
ctcccaactc atccgcagg ttacccctt gaccacgtg tcaatcgct ttc当地ggaa
ccagatttg gcgc当地cccg cagcccccac catcaccacc gtcagtgaaa acgttctgc
tctcacagat cacgggacgc taccgctgcg caacacgc cggaggagtcc agcgagtgac
cattactgac gccagacgc gcacccctgc ctacgtttac aaggccctgg gcatagtc
gccc当地ccgc tcattcgagcc gcactttt agcaagcatg tccatcctt tatc当地ccag
caataacaca ggctggggcc tgcccttccc aagcaagatg tt当地ggggg ccaagaagcg
ctccgaccaa caccaggcgc gcgtgc当地ccgg gcactaccgc ggc当地ccctgg ggc当地ccacaa
acgccc当地ccgc actggcgca ccaccgtcg tgacgc当地ccatc gacgccc当地gg tggaggaggc
gwgcaactac acgccc当地ccgc cgccaccagg gtccacagg gacgccc当地gg ttc当地ggacc
ggtgc当地ccgg gccc当地ccgcgt atgctaaaat gaagagacgg cggaggcgcg tagcacgtcg
ccaccgccc当地ccgc cgaccggcgc ctgccc当地ccca acgc当地ccggcg gggccctgc ttaaccgc
acgtcgaccc ggccgacccgg cggccatgcg ggccgctcga aggctggccg cgggtattgt
caactgtggccccc cccaggctca ggc当地ccggc ggc当地ccggca gcaaggccggg ccattagtgc
tatgactcag ggtc当地ccggg gcaacgtgta tt当地ggg gtcgactcgatgta gccc当地ccctgc
cgtgccc当地ccgt cgccacccgc当地ccccc ccccgccgca ctagattgca agaaaaaaact acttagactc
gtactgttgc atgtatccag cggccglocalccggc ggc当地ccacgaa gctatgtcga agcgcaaaat
caaagaagag atgctccagg tcatcgccg ggagatctat ggccccc当地ccga agaaggaaaga
gcaggattac aagccccc当地ccga agctaaaggcg ggtcaaaaag aaaaagaaaag atgatgatga
tgaacttgc gacgagggtgg aactgctgca cgctaccgc当地ccggc gggtaactgt
gaaagggtcg cgc当地ccaaaac gtgttgc当地ccg accccggcacc accgtactt ttaaccgc当地ccgg
tgagc当地ccgtcc accccgcaccc acaaggc当地ccgt gtatgatgag gtgtacccg localccgg
gctt当地ccggc gccaacgagc gcctc当地ccggg gtttgc当地ccctac gaaaaggc当地ccgg
gctt当地ccggc当地ccgt ccgtccggc当地ccgg agggcaaccc aacacccatgc ctaaaggcc
gcagggtgctg ccccgccctt caccgtccg agaaaaaggc当地ccgc ggc当地ccaaagg
tgacttgc当地ccccc cccaccgtgc agctgatggt accccaaggc当地ccgc cagc当地ccactgg
ggaaaaaaatg accgtggaaac ctgggctgga gccc当地ccggc cgc当地ccggc当地ccgg
ggtggccccc ggacttggc当地ccgt tgccgacttgc当地ccgt ggacglocalcc
cagtattgccc accccgcacag agggcatgga gacacaaaacg tccccc当地ccgtt
ggccggatgccc gccgtgc当地ccgg cggc当地ccgtc当地ccggcc aagaccc
aacggaccccg tggatgttcc gcttccatgc cccccc当地ccggc cccgc当地ccgg
ccggccccc当地ccggc agc当地ccgtac tgccc当地ccgaata tgccct
ccggctatcgt ggctacaccc accccccccag aagacgagc当地ccgc actacc
caactggaaacc cggccccc当地ccggc gtc当地ccgtc当地ccggcc
cagggtggct cgc当地ccggag gcaggaccct ggtgactgcca acagc当地ccgc
catcgtttaa aagccggctt ttgtggtcc tgc当地ccgatatg
tttccccc当地ccgtt ccgggattcc gaggaagaat
cctgacggcc ggc当地ccatgc当地ccgt
gccc当地ccgggtt atccgc当地ccccc
ccggatgtca tccgtggccct tgc当地ccggc当地ccgg
ggaaaaatcaa aaaaaaaaaatg ctggacttcc
gaatggaaaga catcaactt
gaaaactggca agatatcgcc
tgtggaggccg cattaaaaat ttc当地ccgttcc
acagcaggac accccaggatg
tggtagatgg cctggccctt
15060
15120
15180
15240
15300
15360
15420
15480
15540
15600
15660
15720
15780
15840
15900
15960
16020
16080
16140
16200
16260
16320
16380
16440
16500
16560
16620
16680
16740
16800
16860
16920
16980
17040
17100
17160
17220
17280
17340
17400
17460
17520
17580
17640
17700
17760
17820
17880
17940
18000
18060
18120
18180
18240
18300

aaaataagat	taacagtaag	cttgatcccc	gccctccgt	agaggagcct	ccaccggcg	18360
tggagacagt	gtctccagag	gggcgtggcg	aaaagcgtcc	gcgccccgac	agggaaagaaa	18420
ctctgggtac	gcaaataagac	gagcctccct	cgtacgagga	ggcaactaaag	caaggccctgc	18480
ccaccacccg	tcccattcgcg	cccatggcta	ccggagtgct	gggcacagcac	acacccgtaa	18540
cgctggacct	gcctcccccc	gccgacaccc	agcagaaaacc	tgtgtgtcc	ggcccgaccg	18600
ccggttggtt	aaccgttcct	agccgcgcgt	ccctgcgcgg	cgccgcgcagc	ggtccgcgat	18660
cggttgcggcc	cgttagccagt	ggcaactggc	aaagcacact	gaacagcatac	gtgggtctgg	18720
gggttcaatc	cctgaagcgc	cgacgatgt	tctgaatagc	taacgtgtcg	tatgtgtgtc	18780
atgtatgcgt	ccatgtcgcc	gccagaggag	ctgctgagcc	gccgcgcgc	cgctttccaa	18840
gatggctacc	ccttcgtatga	tgccgcagtg	gtcttacatg	cacatctcg	gccaggacgc	18900
ctcggagttac	ctgagccccg	ggctgggtc	gtttgcgcgc	gccaccgaga	cgtacttcag	18960
cctgaataac	aagtttagaa	acccacgggt	ggcgcctacg	cacgacgtga	ccacagacccg	19020
gttccacgggt	ttgacgctgc	ggttcatccc	tgtggaccgt	gaggatactg	cgtactcgta	19080
caaggcgcgg	ttcacccctag	ctgtgggtg	taaccgtgt	ctggacatgg	cttccacgta	19140
ctttgacatc	cgcgccgtgc	tggacagggg	ccctacttt	aagccctact	ctggcactgc	19200
ctacaacgcc	ctggctccca	agggtgcccc	aaatccctgc	gaatgggatg	aagctgctac	19260
tgctttagaa	ataaaacctag	aagaagagga	cgatgacaac	gaagacgaaag	tagacgagca	19320
agctgagcag	caaaaaactc	acgtattttg	gcaggccct	tattttggta	taaatattac	19380
aaaggaggg	attcaaatag	gtgtcgaagg	tcaaaccact	aaatatgcgg	ataaaaacatt	19440
tcaacctgaa	cctcaaatag	gagaatctca	gtgttacgaa	actgaaattt	atcatgcagc	19500
tgggagagtc	ctaaaaaaa	ctaccccaat	gaaaccatgt	tacggttcat	atgaaaacc	19560
cacaaatgaa	aatggagggc	aaggcattct	tgtaaagcaa	caaaatggaa	agctagaaaag	19620
tcaagtggaa	atgcaattt	tctcaactac	tgaggcgtcc	gcaggcaatg	gtgataactt	19680
gactcctaaa	gtggatttgt	acagtgaaga	tgttagatata	gaaaccccgag	acactcatat	19740
tttcttacatg	cccactatta	aggaaggtaa	ctcacgagaa	ctaattggcc	aacaatctat	19800
gccccacagg	cctaattaca	ttgttttag	ggacaatttt	attttgtctaa	tgttaccaa	19860
cagcacgggt	aatatgggt	ttctggcggt	ccaagcatcg	cagttgaatg	ctgtttagaa	19920
tttgcaagac	agaaaacacag	agctttcata	ccagttttg	cttgattcca	ttgggtatag	19980
aaccaggat	ttttctatgt	ggaatcaggc	ttgttacgac	tatgtatccag	atgttagaat	20040
tattgaaaat	catggaaactg	aatgatgaact	tccaaattac	tgctttccac	tgggaggtgt	20100
gattaataca	gagactctt	ccaaggtaaa	acctaaaaaca	ggtcaggaaa	atggatggaa	20160
aaaagatgct	acagaatttt	cagataaaaa	tgaataaaga	gttggaaata	attttgcctat	20220
ggaaatcaat	ctaaatgcca	acctgtggag	aaatttctgt	tactccaaca	tagcgctgt	20280
tttgcggac	aagctaaagt	acagtccttc	caacgtaaaa	atttctgtata	acccaaacac	20340
ctacgactac	atgaaacaage	gagtgggtt	tcccgggtt	gtggactgt	acattaacct	20400
tggagcactc	ttgttccctt	actatatgg	caacgtcaac	ccatttaacc	accaccgca	20460
tgctggctcg	cgcttcccgct	caatgttgc	ggcaatgg	cgctatgtgc	ccttccacat	20520
ccaggtgcct	cagaagtct	ttgcccattaa	aaacccctt	ctccgtccgg	gctcatacac	20580
ctacgagttg	aacttcagga	aggatgtt	catggttctg	cagagctccc	tagaaatgaa	20640
cctaagggtt	gacggagcca	gcattaagtt	tgtatgtt	tgcctttacg	ccacccctt	20700
ccccatggcc	cacaacaccg	cctccacgt	tgaggccat	ctagaaaacg	acaccaacga	20760
ccagtcctt	aacgactatc	tctccgcgc	caacatgtc	taccctatac	ccgccaacgc	20820
taccaacgtg	ccccatatcca	tcccctcccg	caactggcg	gtttccgcg	gttggccctt	20880
cacgcgcctt	aagactaagg	aaaccccatc	actgggtctg	ggctacgacc	cttattacac	20940
ctactctggc	tctataccct	acctagatgg	aaccccttac	ctcaaccaca	cctttaagaa	21000
ggtggccatt	acctttgact	tttctgtc	ctggccttgc	aatgaccg	tgcttacccc	21060
caacgagttt	gaaattaagc	gctcagtt	cggggggg	tacaacgtt	cccaagtgtaa	21120
catgaccaaa	gactgggttcc	ttgtacaaat	gctagtaac	tacaacattt	gttaccagg	21180
tttcttata	ccagagagct	acaaggaccg	catgtactcc	tttttttagaa	actttccagcc	21240
catgagccgt	caggtgggtt	atgataactaa	atacaaggac	taccaacagg	tggcatacct	21300
acaccaacac	aacaactctg	gattttgtt	ctacccgtcc	cccaccatgc	gcaaggacaa	21360
ggccttccct	gctaacttcc	cctatccgc	tataggcaag	accgcgtt	acagcattac	21420
ccagaaaaag	tttcttgc	atcgacccct	ttggcgcate	ccatttccca	gtaaacttat	21480
gtccatgggc	gcactcacag	acctggccca	aaaccccttc	tacgccaact	ccgcccacgc	21540
gctagacat	acttttgg	ttggatccat	ggacgagccc	acccttctt	atgttttgg	21600

tgaaagtcttc gacgtggtcc gtgtgcacccg gccgcaccgc ggctgtcatcg aaaccgtgt
ccttcgcacg cccttctcg ccggcaacgc cacaacataa agaaggcaagc aacatcaaca
acagctgccg ccatgggctc cagtggcag gaactgaaag ctattgtcaa agatcttgg
tgtggccat attttttggg cacctatgac aagcgtttc caggcttgc ttctccacac
aagctcgct gcgcctatgt caatacgccc ggtcgccaga ctgggggcgt acactggatg
gccttgcct ggaacccgc acaaaaaaca tgctacctt ttgagccctt tggctttct
gaccagcgc tcaagcagg taccagg ttagtacgact cactcctgcg ccgtagcgc
attgtttctt cccccgaccc ctgtataacg ctggaaaagt ccacccaaag cgtacagg
cccaactcgg ccgcctgtgg actattctgc tgcatgtttc tccacgcctt tgccaactgg
ccccaaactc ccatggatca caacccccc accatgaccc ttaaccggg acccaactcc
atgctcaaca gtccccagg ttagtacgact cactcctgcg accaggaaca gctctacagc
ttcctggagc gcaactcgcc ctacttccgc agccacagtg cgtagattag gagcggcact
tcttttgc acggaaaaaa catgtaaaaaa taatgtacta gagacacttt caataaaaggc
aaatgctttt atttgtacac tctcggtga ttatttaccc ccaccccttgc cgtctgcgc
gtttaaaaat caaagggtt ctgcccgc acaactatgca tgcgtatgcg ccactggcag
cgatactggt gtttagtgct ccacttaaa tcaggcaca ccatccgcgg cagctcggt
aagtttcc acggcaggct ggcacccatc accaactcggt tttagcaggc gggcggcgt
atcttgaagt cgcagttggg gcctccgc tgcgcgcgcg agttgcgata cacagggtt
cagcactgga acactatcag cggccgggtt tgcacgcgtt ccagcactcg cttgtcggag
atcagatccg cgtccaggc tccgcgtt ctcaggcga acggagtcaa ctttggtagc
tgcctccca aaaaggcgc gtgcccaggc tttgagttgc actcgcaccc tagtggcata
aaaagggtac cgtgcccggt ctggcgtt ggatacagcg cctgcataaa agccttgatc
tgcttaaaag ccacctgac ctttgcgtt tcagagaaga acatggcgc aagacttgcg
gaaaactgtat tggccggaca ggccgcgtcg tgcacgcgc accttgcgtc ggtgttggag
atctgcacca catttcggcc ccacccgtt ttcacgatct tggccttgc agactgctcc
ttcagcgcgc gctgcccgtt ttcgctcg ttcacgatctt caatcactg ctccttattt
atcataatgc ttccgtgtat acacttaaagc tgccttgc ttcagcgcgc ggggtgcgc
cacaacgcgc agccgtggg ctcgtatgc tttgtaggtca cctctgc aaaa cgactgcagg
tacgcctgca ggaatcgc cccatcg tccacgatct tggccttgc agactgctcc
tgcaacccgc ggtgctcctc gttcagccag gtcttcata cggccgcgc agcttcact
tggtcaggca gttagttgaa gttcgcctt agatcggtt ccacgtggta cttgtccatc
agcgcgcgc cagcctccat gccccttgc tccacgacca cgcacgcgc acctcagg
ttcattacccg taatttcaact ttcgcgttcc ctggcgttcc cctcttgc ttcgcgtccgc
ataccacgcg ccactgggtc gtcttcattt agccgcgc ttcacgatctt accttcattt
ccatgctga ttacgcacccg tgggttgc tttgtacgcg cccatcttct
cttcttccct cgctgtccac gattacctt ggtgtatggc ggcgcgtccggg cttggagaa
gggcgtctt tttttttt gggcgcata gccaaatccg cccgcgcaggc cgatggccgc
gggctgggtg tgcgcggcac cagcgcgtt tttgtatggcgtt ctgcgtccgc
atacgcgcgc tcatccgtt ttttggggc gcccggggag gggcgcgcgc cggggacgg
gacgacacgt cctccatgg tgggggacgt cgcgcgcac cgcgtccgc ctcgggggt
gtttcgctt gctcttcc cgcactggcc atttccttcc cctataggca gaaaaagatc
atggagtcag tcgagaagaa ggacagccca accgcgcgc ttcacgatctt caccacgc
tccaccatgc cgcacccgc gcctaccacc tttccgtcg aggacccccc gcttggagg
gaggaagtga ttatcgagca ggaccagg tttgtatggc aagacgcacga ggaccgc
gtaccaacag aggaaaaaaa gcaagaccag gacaacgcac agggcaacgc ggaacaagtc
gggcgggggg acgaaaggca tggcgcactac ctagatgtgg gagacgcacgt gctgttgaag
catctgcacg gccacgtgcgc cattatctgc gacgcgttgc aagacgcacgc cgcacgc
ctcgccatag cggatgtcag cttgcctac gacgcgcacc tattctcacc ggcgc
cccaaacgc aagaaaaacgg ccatgcgag cccaaacccgc gcctcaactt ctacccgc
tttgcgtgc cagagggtct tgccacccat cccatctt tccaaaaactg caagataccc
ctatccgtcc gtgcacccgc cagccgcgc gacaacgcac tggcgtccgc gcaggccgc
gtcataccctg atatcgctc gtcacacggaa gtgcacccaa tcccttgc ggttggacgc
gacgagaagc ggcggccaaa cgctctgc aaggaaaaaca gcaaaaaatga aagtcactt
ggagtgttgg tggaaactcga ggggtacaaac ggcgcgcctag cgcgtactaa acgcacgc
gaggtcaccc actttgccta cccggcactt aacctacccc ccaaggtcat ggcacagtc

atgagtgagc tgatcggtcg ccgtgcgcag cccctggaga gggatgc当地 24960
caaacagagg agggcctacc cgcaatttggc gacgaggcgc tagcgctg gcttcaaact 25020
cgcgacccctg ccgacttggc ggagcgcacgc aaactaatga tggccgc当地 25080
gtggagctt agtgcacatgc gcggttctt gctgaccggg agatgc当地 25140
gaaacatttc actacaccctt tcgacaggc tacgtacgcc aggccctgcaa gatctccaa 25200
gtggagctt gcaacacttgc ctccatccctt ggaattttgc acggaaaaccg ccttgggcaa 25260
aacgtgc当地 attccacgcg caagggcggg ggcgc当地 actacgtccg cgactgc当地 25320
tacttatttc tatgtcacac ctggcagacg gccatggcgg tttggcagca gtgcttggag 25380
gagtgc当地 tcaaggagct gcagaaaaactg ctaaagc当地 acttgc当地 cctatggacg 25440
gccttcaacg agcgc当地 ggccgc当地 ctggcggaca tcattttccc cgaacgc当地 25500
cttaaaaaccc tgcaacacggg tctggc当地 acgc当地 aaagcatgtt gcagaacttt 25560
aggaactt当地 tccttagagcg ctc当地 gaatttgc ttgccc当地 cctgctgtc当地 25620
gactttgtc当地 ccattaaatgc cccgc当地 atgc当地 ccttgc当地 tttgggcaa ctgctacctt 25680
ctgc当地 agc当地 ccaacttgc当地 tctggc当地 tctgac当地 tggaaagacgt gagcgggac 25740
ggtctactgg agtgc当地 tctggc当地 acgc当地 ctatgc当地 cgacccgc当地 cctggglocal 25800
aatttc当地 tgctt当地 acgc当地 aagtcaattt atcggtacct ttgagctgc gggccctcg 25860
cctgacgaaa agtccgc当地 tccggggtt当地 aaactcactc cggggc当地 gacgtcgct 25920
tacccgc当地 aatttttgc当地 tgaggactac cacgc当地 cccggc当地 agatttagtt ctacgaagac 25980
caatccc当地 cgccaaatgc ggagcttacc gcctgc当地 ttaccgc当地 ccacatttcc 26040
ggccaaatttgc aagccatcaa caaagccgc当地 caagagtttgc tgctacgaaa gggacggggg 26100
gtttacttgg accccc当地 cggc当地 gagggg ctc当地 accccaa tccccccgc当地 gccgc当地 26160
tatcagc当地 agccgc当地 ccttgc当地 caggatggc当地 cccaaaaaaa agctgc当地 26220
gccgc当地 cccacggc当地 aggaggacata ctggc当地 caggc当地 gaggggagg 26280
cgaggaggag gaggacatgc tggaaagactg ggagacgc当地 gacgaggac 26340
cgaagagggtg tcagacgaaa caccgc当地 ctc当地 ggccgc当地 cggccccc当地 26400
gaaatcgcc当地 accggtcc当地 gcatggc当地 aacccgc当地 cctcaggc当地 cgccggc当地 26460
gccc当地 ttgc当地 cgacccaaacc gtagatggg当地 caccactggc当地 accaggccg gtaagtccaa 26520
gcagccgc当地 ccttgc当地 aagagcaaca acagc当地 cccaa ggttaccgc当地 catggc当地 26580
gcacaagaac gccatgtt当地 ttgcttgc当地 agactgtggg ggaacacatct ccttc当地 26640
ccgctt当地 ctctaccatc acggc当地 tggc当地 cttccccc当地 aacatccgc当地 attactacc 26700
tcatctctac agcccaactt gcacccggc当地 cagc当地 ggccagcaaca gcagcggc当地 26760
cacagaagca aaggcgaccg gatagcaaga ctctgacaaa gccaagaaa tccacagc当地 26820
cggc当地 cggc当地 aggaggagga gc当地 tggc当地 cggc当地 cggccccc当地 26880
agcttagaaa caggatttt cccactctgt atgctatatt tcaacagacg aggccc当地 aag 26940
aacaagagct gaaaataaaa aacagggtctc tgctgatccct caccgc当地 tgc当地 27000
acaaaagcga agatcactt cggc当地 caccgc当地 tggaaagacgc ggaggctctc ttca 27060
actgc当地 cgc当地 gactt当地 taag gacttagttt ggc当地 cttt当地 tcaaattt当地 ggc当地 27120
tacgtcatct ccagc当地 ggcc当地 caccgc当地 cggc当地 caccctgt cgtc当地 cgc当地 attatgagca 27180
agggaaattcc cacccctac atgtggagtt accagccaca aatgggactt gc当地 ctgggactt 27240
ctgccc当地 aaga ctactcaacc cgaataaaact acatgacgc当地 gggacccccc当地 atgatatccc 27300
gggtcaacgg aatccgc当地 caccgc当地 aaccggc当地 gaatttctt当地 ggaacaggcg gctattacca 27360
ccacacccctc当地 taataaccctt aatcccgta gttggccgc当地 tggccctgt当地 taccaggaaa 27420
gtcccgctcc caccactgtg gtacttccca gagacgc当地 gggc当地 gaaggtt cagatgacta 27480
actcaggggc当地 gc当地 agcttgc当地 ggccgctt当地 gtc当地 acagggtt gggc当地 cgggactt 27540
taactcactt gacaatcaga gggc当地 gaggtt ttc当地 agcttca local cgaacgactcg gtgagctct 27600
cgcttggctt ccgtccggc当地 gggacattt当地 agatcggc当地 cggccggc当地 cttt当地 cattca 27660
cgc当地 cgtc当地 ggcaatccta actctgc当地 cctc当地 tgacgc当地 cggc当地 tctggaggca 27720
ttggaaactt当地 gcaatttt当地 gaggaggttt tgccatcggt ctactt当地 aac cc当地 ttctcg 27780
gacccccc当地 ccactatccg gatcaat当地 ttccat当地 tgc当地 cggc当地 aaggactcg 27840
cgacgc当地 cgtc当地 cggc当地 gagatgc当地 ttaacttgc当地 aggacgacgca actgc当地 cctg 27900
tccactgtc当地 cggc当地 cacaag tgcttgc当地 gcaactccgg tgatgggactt tacttt当地 27960
tgccc当地 gaggtt当地 tcatatcgc当地 ggccggc当地 acggc当地 tccg当地 gttaccgc当地 cagg 28020
ttggccctgtag cctgattcgg gagtttaccgc当地 agc当地 cccctt gctatgtgag cgggacagg 28080
gaccctgtgt tctcactgtg atttgc当地 aactt gtc当地 taacccat tggattacat caagatctt 28140
gttgc当地 catct ctgtgctgag tataataaaat acagaaatta aaatatactg gggctctat 28200

cgccatcctg taaacgcac cgttttcaacc cgcccaagca aaccaaggcg aacccattac
ggtaacttta acatcttcg cttctgttatt tacaacagt tcaacccaga cggaggtagt
ctacgagaga acctctccga gctcagctac tccatcagaa aaaacaccac cctcccttacc
tgccggaaac gtacgagtgc gtcaccggcc gctgcaccac acctaccgccc tgaccgtaaa
ccagacttt tcggacaga cctcaataac tctgttacc agaacaggag gtgagcttag
aaaacccta gggatttagg ccaaaggcgc agctactgtg gggtttatga acaattcaag
caactctacg ggctattcta attcaggtt ctctagaatc ggggttgggg ttattctctg
tcttgttattt ctataacta cttttcttcg ctaaggctcg ccgcctgctg
tgtgcacatt tgcattttt gtcagcttt taaacgctgg gtcgccacc caagatgatt
aggtacataa tcctaggtt actcaccctt gcgtcagccc acggtaccac caaaaagggt
gattttaaagg agccagcctg taatgttaca ttcgcagctg aagctaatga gtgcaccact
cttataaaat gcaccacaga acatgaaaag ctgttattt gccacaaaaaa caaaattggc
aagtatgctg ttatgttat ttggcagcca ggtgacacta cagagtataa tgttacagtt
ttccaggta aaagtataa aactttatg tatactttt cattttatga aatgtgcac
attaccatgt acatgagcaa acagtataaag ttgtggcccc cacaaaattt tttggaaaac
actggactt tctgtgcac tgctatgtc attacagtgc tgccttggg ctgtacccta
ctctatatta aatacAAAAG cagacgcagc ttatttggg aaaaagaaaaat gccttaattt
actaagttac aaagctaattt tcaccactaa ctgcttact cgctgcttgc aaaaacaatt
caaaaagttt gcattataat tagaatagga tttaaaccccc cggcgtcattt cctgctcaat
accattcccc tgaacaattt actctatgtg gatatgtc cagcgtacca accttgaagt
caggcttcctt ggatgtcagc atctgacttt ggccagcacc tggccggg atttgttcca
gtccaactac agcggccac cctaacagag atgaccaaca caaccaacgc ggccggcgt
acgggactt catctaccac aaatacacc ccaagtttctg ctttgcattt taactggat
aacttggca tgggttgcattt ctccatagcg ttatgtttt tattatgtgg
ctcatctgtc gcctaaagcg caaacgcgccc cgaccaccca ttatgttcc catcattgtg
ctacacccaa acaatgttgc aatccataga ttggacggac tggaaacacat gttctttct
cttacagtat gattaaatga gacatgattt ctgcattttt tatattactg accctgttg
cgcttttttgc tgcgtctcc acattggctg cgggttctca catcgaagta gactgcattt
cagccttcac agtctattt ctttacggat ttgttccctt caccgttcc tgcagcctca
tcactgttgtt catcgccctt atccagtgc ttgactgggt ctgtgtgcgc tttgcataatc
tcagacacca tccccagttc agggacagga ctatagctga gtttctttaga attctttaat
tatgaaattt actgtgactt ttctgttgcattt tatttgcacc ctatctgcgt tttgttcccc
gacctccaaag cctcaaagac atatcatg cagatttact cgtatatggg atattccaaag
ttgctacaat gaaaaaaagcg atcttccoga agcctggta tatgcaatca tctctgttat
ggtgttctgc agtaccatct tagcccttagc tatatatccc taccttgaca ttggctggaa
acgaatagat gccatgaacc acccaacttt ccccgccccc gctatgttcc cactgcaaca
agttgttgc ggcggctttc tcccagccaa tcagcctcgc cccacttctc ccaccccccac
tgaaatcagc tactttatc taacaggagg agatgactga cacccttagat cttagaaatgg
acggaatttat tacagagcag cgcctgttag aaagacgcag ggcagcggcc gagcaacagc
gcatgaatca agagctccaa gacatggta acttgcacca gtgaaaagg ggtatcttt
gtctggtaaa gcaggccaaa gtcacctacg acagtaatac caccggacac cgccttagct
acaagttgc aaccaagcg tggaaatttttgg tggcatgtt gggagaaaaag cccattacca
taactcagca ctggtagaa accgaaggct gcattcaactc accttgcattt ggacctgagg
atctctgcac ctttattttt accctgttgcg gtcttacttcc tttactaat
aaaaaaaaat aataaagcat cacttactta aaatcagttt gcaaaatttctt gtccagttt
ttcagcagca cctcccttgc ctcttccctt cttctgttattt gcaacttccctt cctggctgca
aactttctcc acaatctaaa tggaaatgtca gtttccctt gtttcttgc atccgcaccc
actatcttca tgggttgc gatgaagcgc gcaagaccgt ctgaagatac cttcaaccccc
gtgtatccat atgacacggaa aaccggctt ccaactgtgc ctttttttccctt ttttccctt
gtatccccca atgggtttca agagactccc cctgggttac ttttttttgc cttatccgaa
cctctagttt cctccaaatgg catgttgcg ctcaaaaatgg gcaacccgcctt ctctctggac
gaggccggca accttaccc ccaaaaatgtt accactgtga gcccacccctt caaaaaaaacc
aagtcaaaaca taaaaccttggaa aatatctgc cccctccacag ttacctcaga agccctaact
gtggctggcc cccgacccctt aatggtgcg ggcacacac tcaccatgc atcacaggcc
ccgctaaaccg tgcacgactc ctttttttccctt ctttttttccctt ttttccctt

gaaggaaaagc tagccctgca aacatcaggg cccctcacca ccaccgatag cagtaccctt	31560
actatcactg cctcaccccc tctaactact gccactggta gctgggcatt tgacttgaaa	31620
gagcccatt atacacaaaaa tggaaaacta ggactaaagt acggggctcc tttgcattgt	31680
acagacgacc taaacacttt gaccgttagca actgggtccag gtgtgactat taataatact	31740
tccttgcaaa ctaaagttac tggagccctt ggttttggatt cacaaggcaa tatgcaactt	31800
aatgttagcag gaggactaag gattgattct caaaacagac gccttataact tgatgttagt	31860
tatccgtttt atgctcaaaa ccaactaaat ctaagactag gacagggccc tcttttata	31920
aactcagccc acaacttggta tattaactac aacaaggccc ttacttggt tacagcttca	31980
aacaattcca aaaagcttga ggttaaccta agcaactgcca aggggttgat gtttgcgt	32040
acagccatag ccattaatgc aggagatggg cttgaattt gttcacctaa tgccacaaac	32100
acaatcccc tcaaaaacaaa aattggccat ggcctagaat ttgattcaaa caaggctatg	32160
gttccctaaac taggaactgg ctttagttt gacagcacag gtgccattac agtaggaaac	32220
aaaataatgttataaacttcaac tttgtggacc acaccagctc catctcttca ctgttagacta	32280
aatcgagaga aagatgttacttcaacttgcgtttaacaa aatgtggcag tcaaataactt	32340
gctacagttt cagttttggc tggtaaaggc agtttgctc caatatctgg aacagtccaa	32400
agtgcctatc ttattataag atttgacgaa aatggagtgc tactaaacaa ttccctctg	32460
gaccggaaat attggaaactt tagaaatggaa gatcttactg aaggcacagc ctatacaaac	32520
gctgttggat ttatgcctaa cctatcagct tatccaaaat ctacacggtaa aactgccaa	32580
agtaacatgg tcagtcatttggc ttacttacgg aagacaaaaa ctaaacctgt aacactaacc	32640
attacactaa acggtacaca ggaaacagga gacacaactc caagtgcata ctctatgtca	32700
ttttcatggg actggcttgg ccacaactac attaatggaaa tatttgcac atccctttac	32760
actttttcat acattgccc aagataaaaga atcgtttgtt ttagtttca acgtgtttat	32820
ttttcaatgg cagaaaaattt caagtcattt ttcatttcgt agtatacgcc caccaccaca	32880
tagcttatac agatcaccgtt accttaatca aactcacaga acccttagtat tcaacctgccc	32940
acctccctcc caacacacag agtacacagt cctttctcc cggttggccct taaaagcat	33000
catacatgg gtaacagaca tattcttagg tggtagatcc cacacggttt cctgtcgagc	33060
caaacgctca tcagtgatataataaaactc cccgggcagc tcacttaagt tcatgtcgct	33120
gtccagctgc tgagccacag gctgctgtcc aacttgcggg tgcttaacgg gcccgcagg	33180
agaagtccac gcctacatgg gggtagagtc ataatgtgc atcaggatag ggcgttgg	33240
ctgcagcagc ggcgcaataa actgctgccc cggccgtcc gtcctgcagg aatacaacat	33300
ggcagttggc tcctcagcga tgattcgcac cggccgcagc ataaggcgcc ttgtccctcg	33360
ggcacagcag cgcaccttga tctactttaa atcagcacag taactgcagc acagcaccc	33420
aatattgttc aaaatcccc acgtcaaggc gctgtatcca aagctcatgg cggggaccac	33480
agaacccacg tggccatcat accacaagcg caggttagatt aagtggcagc ccctcataaa	33540
cacgtggac ataaacattt cctttttgg catgtttaa ttcaccaccc cccgttacca	33600
tataaaccctc tgattaaaca tggcgccatc caccaccatc ctaaaccaggc tggccaaac	33660
ctggcccgcc gctataact gcagggaaacc gggactggaa caatgacagt ggagagccca	33720
ggactcgtaa ccatggatca tcatgtctcgatcatatca atgttggcag aacacaggca	33780
cacgtgcata cacttcctca gattacaag ctcctccgc gtttagaacc tatcccagg	33840
aacaacccat tcctgaatca gcgttaatcc cacactgcag ggaagaccc gcacgttaact	33900
cacgttgcattt attgtcaag tggtagatcc gggcagcagc ggatgtatcc ccagtatgg	33960
agcgcggggtt tctgtctcaa aaggaggtt acgatccctt ctgtacggag tgcggcggaga	34020
caaccggat cgtgttggc gtatgttcat gccaaatggg acggccggacg tagtcatatt	34080
tcctgaagca aaaccagggtt cggggcgtgac aaacagatct gctgtcccg tctcggcgt	34140
tagatcgctc tggtagatgg tggtagatata tccactctt caaagcatcc aggcgcggcc	34200
tggcttcggg ttctatgtaa actccttcat gggccgtcc cctgataaca tccaccaccc	34260
cagaataagc cacacccagc caacctacac attcggttcc cgagtacac acggggagg	34320
cgggaaagagc tggagaacc atgtttttttt ttttatttca aagattatc caaaacccca	34380
aatgaagat ctattaagtgc aacgcgtcc cctccgggtt cgtggcaaa ctctacagcc	34440
aaagaacaga taatggcatt tggtagatgt tgcacaatgg ctggccaaaag gcaaacggcc	34500
ctcacgttca agtggacgtt aaggctaaac ccttcagggtt gaatctccct tataaacatt	34560
ccagcacctt caaccatgccc caaataattt tcatctcgcc accttcttca tatactctt	34620
agcaaatccc gaatattaag tccggccattt gtaaaaaatct gctccagagc gcccctccacc	34680
ttcagcctca agcagcgaat catgattgca aaaaatttccagg ttccctcacag acctgtataa	34740
gattcaaaaag cggaacattt aaaaaataac cgcgatcccg taggtccctt cgccggcca	34800

gctgaacata atcgtgcagg tctgcacgga ccagcgccg cacttccccg ccaggaacct	34860
tgacaaaaga acccacactg attatgacac gcatactcg agctatgcta accagcgtag	34920
ccccgatgt agctttgtt catggcgcc gatataaaat gcaagggtct gctaaaaaaa	34980
tcaggcaag cctcgccaa aaaagaaagc acatcgtagt catgctcatg cagataaagg	35040
caggtaagct ccggaccac cacagaaaaa gacaccattt ttctctcaaa catgtctcg	35100
ggtttctgca taaacacaaa ataaaataac aaaaaaacat ttaaacattt gaagcctgtc	35160
ttacaacagg aaaaacaacc cttataagca taagacggac tacggccatg ccggcgtagc	35220
cgtaaaaaaaaa ctggtcaccg tgattaaaaa gcaccaccg cagctcctcg gtcatgtccg	35280
gagtctataat gtaagactcg gtaaacacat caggttgatt catcggtcg tgctaaaaag	35340
cgaccgaaat agccggggg aatacatacc cgccggcgtt gagacaacat tacagcccc	35400
ataggaggtt taacaaaatt ataggagag aaaaacacat aaacacctga aaaacctcc	35460
tgcctaggca aaatagcacc ctcccgcctc agaacaacat acagcgctc acagcgccag	35520
cctaacagtc agccttacca gtaaaaaaga aaaccttta aaaaaacacc actcgacacg	35580
gcaccagctc aatcgtcac agtgtaaaaa agggccaagt gcagagcgag tatatatagg	35640
actaaaaaat gacgtaacgg ttaaagtcca caaaaaacac ccagaaaaacc gcacgccaac	35700
ctacggccag aaacgaaagc caaaaaaccc acaacttctt caaatcgta cttccgtttt	35760
cccacgttac gtaacttccc attttaaagaa aactacaatt cccaacacat acaagttact	35820
ccggccctaaa acctacgtca cccggcccg tccccacgccc cggccacgtt cacaactcc	35880
acccctcat tatcatattt gttcaatcc aaaataaggt atattattga tgatg	35935

<210> 9

<211> 35935

<212> DNA

<213> Adenovirus serotype 5

<400> 9

catcatcaat aatacacctt attttgattt gaagccaata tgataatgag ggggtggagt	60
ttgtgacgtg gcgcggggcg tgggaacggg gcgggtgacg tagtagtgtg gcggaagtgt	120
gatgttgc当地 gtgtggcgga acacatgtaa gcgcggatg tgccaaaatg gacgttttg	180
gtgtgc当地 gtgtacacag gaagtgc当地 ttttc当地 gtc当地 gttttaggc当地 gatgtttag	240
taaatttggg cgttaaccgag taagatttgg ccattttc当地 ggaaaaactg aataagagga	300
agtgaatctt gaataatttt gtgttactca tagcgc当地 tattttgtctt gggccgccc	360
gactttgacc gtttacgtgg agactcgccc aggtt当地 ctc当地 gttt当地 ttcccgcttc	420
cgggtcaaaat ttggc当地 attttatag tcagctgacg tttttttt当地 ttataaccgg	480
tgagttccctt aagaggccac ttttgc当地 cagcgacggatg agtttctcc tccgagccgc	540
tccgacaccg ggactgaaaa tgagacatata tatctgccc acggatgatg ttaccgaaga	600
aatggccccc agtcttttgg accagctgtat cgaagaggta ctggctgata atcttccacc	660
tccttagccat ttgaaccac ctacccttca cgaactgtat gatttttagacg tgacggcccc	720
cgaagatccc aacgggggggg cgggttccgc gattttccc gactctgtaa ttttttccgg	780
gcagggaggatttacttcc ctttttccgc gtttccgg agccgc当地	840
cctttccccc cagcccgagc agccggagca gagaggctt ggtccgg gtttccgg	900
ccttgc当地 gaggtgatcg atcttacgtt ccacggatg ggttccggcc acgtgacga	960
cgaggatgaa ggggggagg agtttgc当地 agattatgtg gagcaccggc ggcacgggtt	1020
caggtcttgc当地 ctttttccgc gtttccgg agccgc当地	1080
ctatatgagg acctgtggca ttgttgc当地 cagtaatgtaa aatattatggg cagttgggtga	1140
tagagtgggtagtgggatgtt ggttccggcc gtttccggcc acgtgacga	1200
gaattttgtt ttgtgatattt ttttttgc当地 cctgtgtctt aacctgagcc tgagccccc	1260
ccagaaccgg agcctgc当地 accttccgc cgttccctt当地 tgccgc当地 tatcctgaga	1320
cgttccgc当地 caccgtgtcc tagagaatgc aatagtagta cggatagctg tgactccgg	1380
ccttcttccat caccgtgtcc gtttccggcc ttttttccat taaaccaggat	1440
gccgtgagatggggcc ttttttccgc gtttccggcc ttttttccat taaaccaggat	1500
ccttccgc当地 ctttccgc当地 gaggttccggcc ctttccggcc ttttttccat taaaccaggat	1560
ttgcgtgtt ggttccggcc ttttttccat taaaccaggat	1620
gagataatgtt ttaacttgc当地 ttttttccat taaaccaggat	1680
cgccgtggcc ttttttccat taaaccaggat	1740

ttttctgctg	tgcgttaactt	gctggaacag	agctctaaca	gtacctcttg	gttttgagg	1800
tttctgtgg	gctcatccca	ggcaaagtta	gtctcgagaa	ttaaggagga	ttacaagtgg	1860
gaatttgaag	agctttgaa	atcctgtgg	gagcttttg	attctttgaa	tctgggtcac	1920
cagggcgctt	tccaagagaa	ggtcatcaag	actttgatt	tttccacacc	ggggcgcgct	1980
gcggctgctg	ttgcttttt	gagtttata	aaggataaaat	ggagcgaaga	aaccatctg	2040
agcggggggt	acctgctgga	tttctggcc	atgcatctgt	ggagagcggt	tgtgagacac	2100
aagaatccc	tgctactgtt	gtcttccgtc	cgccccggca	taataccgac	ggaggaggcag	2160
cagcagcagc	aggaggaagc	caggcggcgg	cggcaggagc	agagcccatg	gaacccgaga	2220
gccggccctgg	accctcgga	atgaatgtt	tacaggctgc	tgaactgtat	ccagaactga	2280
gacgcattt	gacaattaca	gaggatgggc	aggggcataa	gggggttaaag	agggagcggg	2340
gggcttgtga	ggctacagag	gaggcttagga	atctagctt	tagcttaatg	accagacacc	2400
gtcctgagtg	tattacttt	caacagatca	aggataattg	cgctaattag	cttgatctgc	2460
tggcgcagaa	gtattccata	gaggcgtca	ccacttactg	gctgcagcca	ggggatgatt	2520
ttgaggagc	tattaggta	tatgcaaagg	tggcacttag	gccagatgc	aagtacaaga	2580
tcagcaaact	tgtaaatatac	aggaattttg	gctacatcc	tggAACGGG	gccgagggtgg	2640
agatagatac	ggaggatagg	gtggcctta	gatgttagcat	gataaaatatg	tggccggggg	2700
tgcttggcat	ggacgggggt	tttattatga	atgtaaagg	tactggcccc	aattttagcg	2760
gtacggttt	cctggccaat	accacccctt	tcctacacgg	tgtaaagcttc	tatgggttta	2820
acaatacccg	tgtggaaagcc	tggaccgat	taagggttcg	gggctgtgcc	ttttactgct	2880
gctggaaagg	ggtgtgtgt	cggcccaaaa	gcagggttcc	aattaagaaa	tgcctcttg	2940
aaagggtgtac	cttgggttac	ctgtctgagg	gtaactccag	ggtgcgcccac	aatgtggcct	3000
ccgactgtgg	ttgcttcatg	ctagtaaaa	gctgttgtgt	gattaagcat	aacatggtat	3060
gtggcaactcg	cgaggacagg	gcctctcaga	tgctgacctg	ctcgacggc	aactgtcacc	3120
tgctgaagac	cattcacgt	gccagccact	ctcgcaaggc	ctggccagt	ttttagcata	3180
acatactgac	ccgctgttcc	ttgcatttgg	gtaacaggag	gggggtgttc	ctaccttacc	3240
aatgcaatt	gagtacact	agatattgc	ttgagccga	gagcatgtcc	aaggtaacc	3300
tgaacgggg	gtttgacatg	accatgaaga	tctggaaggt	gctgagggtac	gatgagaccc	3360
gcaccagg	cagaccctgc	gagtgtggc	gtaaacatat	taggaaccag	cctgtgtatc	3420
tggatgtac	cgaggagctg	aggcccgatc	acttggtgct	ggcctgcacc	cgcgctgagt	3480
ttggctctag	cgatgaagat	acagattgag	gtactgaaat	gtgtggcgt	ggcttaaggg	3540
ttggaaagaa	tatataaggt	gggggtctta	tgtagtttg	tatctgtttt	gcagcagccg	3600
ccggcccat	gagcaccaac	tcgtttgat	gaagcattgt	gagctcatat	ttgacaacgc	3660
gcatgcccc	atggccggg	gtgcgtcaga	atgtatgtgg	ctccagcatt	gatggtcgcc	3720
ccgtcctgcc	cgcaactct	actacccct	cctacgagac	cgtgtctgga	acgcccgttgg	3780
agactgcagc	ctccggcc	gttcagccg	ctgcagccac	cgccccgggg	atttgtactg	3840
actttgctt	cctgagcccg	tttgcagca	gtgcagctt	ccgttcatcc	gcccgcgtat	3900
acaagttgac	ggcttcttt	gcacaattgg	attcttgc	ccgggaactt	aatgtcgat	3960
ctcagcagct	gttggatctg	cgccagcagg	tttctggccct	gaaggcttcc	tccctccca	4020
atgcggttt	aaacataaaat	aaaaaaccag	actctgttg	gattttggatc	aagcaagtgt	4080
tttgctgtct	ttatttaggg	gttttgcgc	cgcggtaggc	ccgggaccag	cggtctcggt	4140
cgttgggg	cctgtgtatt	ttttccagga	cgtgtttaag	gtgactctgg	atgttcagat	4200
acatgggcat	aagccgtct	ctgggggtgg	ggttagcacca	ctgcagagct	tcatgtcg	4260
gggtgggtt	gtagatgatc	cagtcgtagc	aggagcgctg	ggcgtgggtc	ctaaaaatgt	4320
cttcagtag	caagtcgtt	gccaggggca	ggcccttgg	gtaagtgttt	acaaaggcggt	4380
taagctgg	tgggtgcata	cgtggggata	tgagatgc	cttggactgt	attttttaggt	4440
tggctatgtt	cccagccata	tccctccggg	gattcatgtt	gtgcagaacc	accagcacag	4500
tgtatccgt	gcacttggg	atttgcata	gtagcttgc	aggaaatgcg	tggaaagaact	4560
tggagaccc	tttgcaccc	ccaagat	ccatgcattc	gtccataatg	atggcaatgg	4620
gcccacgggc	ggcgccctgg	gcaagat	ttctggatc	actaacgtca	tagttgttt	4680
ccaggatgag	atcgtcatag	gccat	caaagcgccg	gcccgggggt	ccagactgcg	4740
gtataatgtt	tccatccggc	ccaggggcg	agttaccctc	acagatttgc	atttccacg	4800
ctttgagg	agatgggggg	atcatgtct	cctgcggggc	gatgaagaaa	acggtttccg	4860
gggttaggg	gatcagctgg	gaagaaagca	ggttcttgc	cagctgcgac	ttaccgcagc	4920
cggtggggcc	gtaaatcaca	cctattacc	ggtgcactg	gtatgttaaga	gagctgcagc	4980
tgccgtatc	cctgagcagg	ggggccactt	cgtaagcat	gtccctgact	cgcattttt	5040

ccctgaccaa atccgccaga aggccgtcg cgcggcga tagcagtct tgcaaggaag	5100
caaaggaaaa caacggttt agaccgtcg ccgttaggc gctttgagc gtttacccaa	5160
gcagttccag gcggccccac agctcggtca cctgcgtcac ggcatctcg tccagcatat	5220
ctccctcggtt cgcgggttgg ggccggcttc gctgtacggc agtagtcggt gctcgccag	5280
acggggccagg gtcatgtctt tccacggggc cagggtcctc gtcagcgtag tctgggtcac	5340
ggtaagggg tgcgtccgg gctgcgcgt ggccagggtg cgcttgaggc tggccctgct	5400
gggtgcgtaa cgctgcgggtt ctgcgcgtt cgcgtccggc agtagtcatt tgaccatgg	5460
gtcatagtc agcccccggc cggcggtggc ctggcgcc acgttgcctt tggaggaggc	5520
gcccgcacgg gggcagtgcgactttttag ggcgttagagc ttggcgccgaa gaaataccga	5580
ttccggggag taggcattccg cggccgcaggc cccgcacgcg gtcgtcgatt ccacgagcca	5640
ggtgagctt ggcgttcgg gtcggaaaac cagggttccc ccatgtttt tgatgcgtt	5700
cttacctctg gtttccatga gccgtgtcc acgctcggtt acggaaaaggc tgtccgttc	5760
cccgatataca gacttgagag ggctgtcctc gagcgggtt ccgcgttcctt cctcgatata	5820
aaactcgacg cacttgaga caaaggctcg cgtccaggcc acgacgaaagg aggctaagt	5880
ggaggggtag cggtcggtt ccactagggg gtccactcg tccagggtgt gaagacacat	5940
gtcgccctct tcggcatcaa ggaagggtat tgggtttagt gtgtaggcca cgtgacccgg	6000
tgttccctgaa gggggctat aaaagggggtt gggggcggt tcgttcctc ac tctttccgc	6060
atcgctgtct gcgaggggcca gctgttgggg tgagtactcc ctctgaaaag cgggcattgac	6120
ttctcgctca agattgtca gttccaaaaa cgaggaggat ttgatattca cctggcccg	6180
ggtgatgcct ttgagggtgg ccgcattcat ctggtcagaa aagacaatct ttttgttgc	6240
aagcttgggt gcaaacgacc ctagaggggc gttggacagc aacttggcga tggagcgcag	6300
ggtttgggtt ttgtcgctat cggccgcgtt cttggcccg atgtttagct gcacgtattc	6360
gcgcgcacgg caccgcatt cggggaaagac ggtggcgc tcgtccggca ccagggtgcac	6420
gcgcacccgg cggttgca ggggtgacaag gtcaacgcgtt gttggctaccc ctccgcgt	6480
gcgcgttgcgtt gtccagcaga ggcggccgc cttgcgcgag cagaatggcg ttaggggtc	6540
tagctgcgtc tgcgtccggg ggtctcgctc cacggtaaag accccgggca gcaggcgc	6600
gtcgaagtag tctatcttc atccttgc caa gtctagcgc tgcgtccatg cgccggccgc	6660
aagcgcgcgc tcgtatgggt tggatgatata ggttggacgc tgaagacgt tgaagctggc	6720
ggcgtacatg ccgcaaatgt cgtaaacgtt gagggtctt ctgagtttcc caagatatgt	6780
aggtagcat cttccaccgc ggtatgcgttgc ggcacgtaa tgcgtatagtt cgtgcgaggg	6840
agcgaggagg tcgggaccga gttgtctacg ggcggctgc tctgctcgga agactatctg	6900
cctgaagatg gcatgtgagt tggatgatata ggttggacgc tgaagacgt tgaagctggc	6960
gtctgtgaga cctaccgcgt cacgcacgaa ggaggcgttag gatgcgcga gcttgggtac	7020
cagctcgccg gtgacctgca cgtctaggggc gcaagtatcc aggtttccct tggatgtgtc	7080
atacttatcc tggccctttt tttccacag ctgcgggttgg aggacaaact ctgcgggtc	7140
tttccagttac tcttggatcg gaaaccgcgc ggcctccgaa cgtaagagc ctgcgtatgt	7200
gaactgggtt acggcctgggt aggccgcagca tccctttctt acgggttagcg cgtatgcctg	7260
cgccggccttc cggagcgagg tgggggttagt gcaaaagggtt tccctgacca tgactttgag	7320
gtactggat ttgaagtcg tgcgtcgca tccggccgtc tcccaagagca aaaaggccgt	7380
gcgccttttgc gaaacgcggat ttggcaggggc gaagggtgaca tgcgttgaaga gatctttcc	7440
cgccgcgaggc ataaagtgcgtt gatgtatgcg gaagggtccc ggcacctcg aacgggtgtt	7500
aattacctgg gcccgcgagca cgtatcgatca aaagccgtt atgttgcgc ccacaatgt	7560
aagttccaag aagcgcggga tggccctgtt ggaaggcaat ttttaagtt cctcgtaggt	7620
gagctttca gggggagctga gcccgtgtc tggccatggc cagtctgca gatgggggtt	7680
ggaagcgcacg aatgagctcc acaggtcaccg ggccatttgc atttgcagg ggtcgcaaaa	7740
ggtcctaaac tggcgaccta tggccattttt ttctgggttgc atgcgtatgtt aggtaaagcgg	7800
gtcttggccc cagcggtccc atccaagggtt cgcggcttgc tgcgtcgccg cagtcactag	7860
aggctcatct cccgcgaact tcatgaccag catgaaggcc acgagctgtc tcccaaggc	7920
ccccatccaa gtataggatctt ctacatcgatca ggtacaaag agacgcgtcg tgcgaggatg	7980
cgagccgatc gggaaagact gatctcccg ccaccaattt gagggtgtgc tattgtatgt	8040
gtgaaaggtag aagtccctgc gacggggccga acactcgatc tggctttgtt aaaaacgtgc	8100
gcagttactgg cagcggtgcgacttccatgc acgttgcacg aggttgcacct gacgaccgc	8160
cacaaggaaag cagagtggga atttgagccctt ctcgcgttgc ggtttggctt ggtggcttc	8220
tacttcggctt gcttgcctt gaccgtctgg ctgcgtcgagg ggagttacgg tggatcgac	8280
caccacgcgcg cgcgcggccca aagtcccgat gtcgcgcgcg ggcggcgcg gcttgcgtac	8340

aacatcgcbc agatgggagc tgtccatggt ctggagctcc cgccggcgtca ggtcaggcgg 8400
 gagctcctgc aggttacct cgcataagac ggtcaggcgg cgggctagat ccaggtata 8460
 cctaatttcc agggctgggt tggtgccggc gtcgatggct tgcagaaggc cgcattcccg 8520
 cggcgcact acggtaaccgc gcggcggggcg gtgggcggcg ggggtgtctt tgatgtatgc 8580
 atctaaaagc ggtgacggcg gcgagcccc ggaggttaggg ggggtctccgg acccgccgg 8640
 agagggggca ggggcacgtc ggcgcggcgc gccggcagga gctgggtctg cgcgcgtt 8700
 ttgctggcga acgcgacgac gccggcggtt atcttcgtaa tctggcgcct ctgcgtgaag 8760
 acgacggggc cggtgagctt gagcctgtca cagaatcaat ttgcgtgtcg 8820
 ttgacggcg cctggcgaa aatcttcgtc acgttctctg agttgtctt ataggcgatc 8880
 tcggccatga actgtcgtat ctcttcgttcc tggagatctc cgcgtccggc tcgcgtccacg 8940
 gtggcggcga ggtcgttggaa atgcggggc atgagctgcg agaaggcg 9000
 tcgttccaga cgcggctgtca gaccacgccc ctttcggcat cgcggcgcg catgaccacc 9060
 tgcgcgagat tgagctccac gtgcggggcg aagacggcg 9120
 aggttagtta ggggtgtggc ggtgttctt gccacgaaga agtacataac ccagcgtcgc 9180
 aacgtggatt cgttgatata ccccaaggcc tcaaggcgct ccatggcctc gtagaagtcc 9240
 acggcgaagt tgaaaaactg ggagttgcgc gccgacacgg ttaacttcctc ctccagaaga 9300
 cggatgagct cggcgcacgt gtcgcgcacc tgcgcgtcaa aggctacagg ggcctctt 9360
 tcttccttcaa ttccttcctt cataaggggc tcccccttctt cttttcttgg cggcggttgg 9420
 ggagggggga cacggcgccg acgacggcgc accgggaggc ggtcgcacaaa ggcgtcgatc 9480
 atctcccccgc ggcgcacggcg catggtctcg gtgcgcgcg ggccttctc gggggggcgc 9540
 agtttggaaaga cgcgcggcg catgtcccg ttagtgggtt gccccgggct gccatgcggc 9600
 agggataacgg cgctaacatgc gcatctcaac aatttgttgc taggtactcc gccgcggagg 9660
 gacctgagcg agtccgcattt gacccggatcg gaaaacctctt cgagaaaaggc gtctaaccag 9720
 tcacagtgcg aagtaggct gacgcacgtt gccggcgccg gccggcgccg gtccgggtt 9780
 tttctggcg aggtgtctgt gatgtatgtaa ttaaagtagg cggctttagt acggcgatg 9840
 gtcgacagaaa gcaccatgtc cttgggtccg gcctgtcaa tgccgcaggcg gtcggccatg 9900
 cccccaggctt cggtttgaca tccggcgagg tctttgtatg agtcttgcattt gggctttctt 9960
 accggcactt cttcttcctt ttccttcgtt cctgcatttc ttgcatttat cgctgcggcg 10020
 gccggcgaggat ttggccgtat gtggcgccctt cttcccttcca tgcgtgtgac cccgaagccc 10080
 ctcatcggtt gaaggcaggcc taggtcgccg acaacgcgtt cggctaatat ggcctgtgc 10140
 acctcggtt gggtagactg gaagtcattt atgtccacaa agcgggttgc tgcgcggctg 10200
 ttgatgggtt aagtgcgtt ggccataacg gaccagttaa cggcttgcgtt acccggttgc 10260
 gagagcttgg tgcgttgcgtt acgcgtatggt acaacacgtt gtcgttgc 10320
 gtccgcacca ggtacttggta tccccaccaaa aagtgcggcg gccggcttggcg gtagagggc 10380
 cagcgtaggg tggccggggc tccggggcgcc agatcttcca acataaggcg atgatatccg 10440
 tagatgtacc tggacatcca ggtgatgcgc gccggcggttgg tggaggcgccg cggaaagtcc 10500
 cggacgcgtt tccagatgtt ggcgcacgcgc aaaaagtgtt ccatggtgcgg gacgctctgg 10560
 ccggtcaggc ggcgcacatc gttgcgttcc tagaccgtgc aaaaggagag cctgtaaacgc 10620
 ggcactcttc cgtggcttgg tggataaattt cgcacggatc tcatggcgatc cgaccgggtt 10680
 tcgagccccg tatccggccg tccggccgtat tccatgcgtt taccggccgcgtt gtcgtcaacc 10740
 cagggtgtcg acgtcagaca acggggggat gtccttttgc gtttgccttcc aggcgcggcg 10800
 gtcgtgtgcg tagcttttttggccacttgcg cgcgcgcgcgtt gtaagcggtt aggctggaaa 10860
 gcgaaagcat taagtggctt gtccttgcgat gccggagggtt tattttccaa ggggttgcgtt 10920
 gccggacccc cgggtcgatgtt ctcggaccgg ccggactgcg gcaacggggg gtttgccttcc 10980
 cccgtcatgtca agaccccgct tgcataattcc tccggaaaca gggacgagcc ctttttttgc 11040
 ttttcccaaga tgcatccgtt gtcgttgcgtt atgcgcggccctt ctcctcagca gccgcacatc 11100
 caagagcagc ggcacatgc caggccaccc tcccccttcc tcaccgcgtc aggaggggcg 11160
 acatccgcgg ttgacgcggc agcagatgtt gattacaac ccccgccggcg ccggggcccg 11220
 cactacctgg acttggatgg gggcgaggccg ctggcgccgc taggagcgcc tccttcgttgc 11280
 cggtaacccaa ggggtcgatgtt gaaacgtgtatgc acgcgttgcgg cgtacgtgc gccgcacac 11340
 ctgtttcgatc accgcgcggg agaggagccg gaggagatgc gggatgcggaa gttccacgc 11400
 gggcgccgatc tgcggcatgg cctgtatgcg gacgggttgc tgcgcgttgc gggactttgc 11460
 cccgcacgcgc gaaacccggat tagtcccgccg cgcgcacacg tggcgccgcg cgcacatgc 11520
 accgcataacg agcagacgcgtt gaaaccaggat attaactttc aaaaagctt taacaaccac 11580
 gtgcgtacgc ttgtggcgccg cgaggaggat gctataggac tgatgcattt gttggactt 11640

gtaagcgcgc tggagcaaaa cccaaatagc aagccgctca tggcgcagct gttccttata 11700
 gtgcagcaca gcagggacaa cgaggcattc agggatgcgc tgctaaacat agtagagccc 11760
 gagggcccgct ggctgctcgta tttgataaaac atcctgcaga gcatagtggt gcaggagcgc 11820
 agcttgagcc tggctgacaa ggtggccgccc atcaactatt ccatgcttag cctggggcaag 11880
 ttttacgccc gcaagatata ccataccccct tacgttccca tagacaagga gttaaagatc 11940
 gaggggtttc acatgcgcatt ggcgcgtgaag gtgccttaccc ttagcgcacga cctggggcgtt 12000
 tatcgcaacg agcgcattcca caaggccgtg agcgtgagcc ggcggccgcga gctcagcgac 12060
 cgcgagctga tgcacagccct gcaaaggccc ctggctggca cggcgcgtgg cgatagagag 12120
 gcccgcgtact ttgttgcgc gggcgctgac ctgcgcgtgg ccccaagccg acgcgcctg 12180
 gagggcagctg gggccggacc tgggctggcg gtggcaccgg cgcgcgtgttgg caacgtccgc 12240
 ggcgtggagg aatatgacga ggacgtatgag tacgagccag aggacggcga gtactaagcg 12300
 gtgtatgtttc tgatcagatg atgcaagacg caacgcaccc ggcgggtgcgg gccgcgtc 12360
 agagccagcc gtccggcctt aactccacgg acgactggcg ccaggtcatg gaccgcattca 12420
 tgcgcgtgac tgcgcgcatt cctgacgcgt tccggcagca gccgcaggcc aaccggctct 12480
 ccgcatttct ggaagcggtg gtccggcgc ggcgcatttcc cacgcacgag aagggtgtgg 12540
 cgatcgtaaa cgcgcgtggcc gaaaacagggg ccattccggcc cgacgaggcc ggcctgttct 12600
 acgacgcgcgt gcttcagcgc gtggctcggtt acaacagccg caacgtgcag accaacctgg 12660
 accggcgtgtt gggggatgtg cgcgaggccg tggcgcagcg tgagcgcgcg cagcagcagg 12720
 gcaacctggg ctccatggtt gcactaaacg ccttccttag tacacagccg gccaacgtgc 12780
 cgcggggaca ggaggactac accaactttg tgagcgcact gcccgttatg gtgactgaga 12840
 caccgc当地 tgaggtgtac cagtctgggc cagactattt ttccagacc agtagacaag 12900
 gcctgcagac cgtaaacctg agccaggctt tcaaaaaactt gcaggggtgttgggggtgc 12960
 gggctccac aggcgcgcg ggcgcgtgtt ctagttgtt gacgcggccac tcgcgcctgt 13020
 tgctgcgtct aatagcgcctt ttcacggaca gtggcgcgtt gtccggggac acatacttag 13080
 gtcacttgcgt gacactgtac cgcgcggccca taggtcaggc gcacgtggac gacatactt 13140
 tccaggagat tacaagtgtc agccgcgcgc tggggcagga ggacacgggc agcctggagg 13200
 caaccctaaa ctacgtgtt accaaccggc ggcagaagat cccctcggtt cacagttaa 13260
 acagcgagga ggagcgcatt ttgcgttgcg tgcagcagg cgtgacgtt aacctgtatgc 13320
 ggcgcgggtt aacgcgcgcg gtggcgtgtt acatgacgcg ggcgcacatg gaaccggca 13380
 tgtatgcctt aaaccggccg ttatcaacc gcctaatttgc ctacttgcgtt cgcgcggccg 13440
 ccgtgaaccc cgagtatttcc accaatgcca tcttgcatttcc gcaactggctt ccccccctg 13500
 gtttctacac cgggggatttcc gaggttgcggc agggtaacga tgatccctc tgggacgaca 13560
 tagacgcacag cgtgtttcc cgcgcacccgc agaccctgtt agagttgcaaa cagcgcgcgc 13620
 aggcaaggccg ggcgcgtgcga aaggaaagct tccgcaggcc aagcagcttgc tccgatctt 13680
 ggcgcgtggc cccgcgggtca gatgttagtgc gcccattttcc aagcttgcata gggctcttta 13740
 ccagcactcg caccacccgc cgcgcgttgc tggcgcaggaa ggagtaccta aacaactcgc 13800
 tgctgcagcc gcagcgcgcgaaa aaaaacttgc ctccggcatt tcccaacaac gggatagaga 13860
 gcctagtggaa caagatgagt agatggaaaga cgtacgcgcgaa ggacgcacagg gacgtggcc 13920
 gcccgcgcgc gcccacccgt cgtcaaaaggc acgaccgtca gccccggctt gttggggagg 13980
 acgatgactc ggcagacgc acgacgcgttcc tggattttggg agggagtggc aacccttttgc 14040
 cgcacccgttcc ccccgaggctg gggagaatgt ttaaaaaaaaaaa aaaaagcatg atgcaaaaata 14100
 aaaaactcac caaggccatg gcaccgcgcg tgggtttct tttatccccc ttagtatgc 14160
 ggcgcgcggc atgtatgagg aaggcttcc tcccttctac gagagtgtgg tgacgcgcgc 14220
 gcccaggccg cggccgttgc gtttccctt cgtatgttcc ctggaccgcgc cgtttgtgcc 14280
 tccgcgggtac ctgcggccta cccggggggag aaacagcatc cgttactctg agttggcacc 14340
 cctattcgcac accacccgtg tttatgttgc ttggatccccc ttagtatgc 14400
 gaactaccag aacgaccaca gcaacttttgc gaccacggc attcaaaaaca atgactacag 14460
 cccggggggag gcaagcacac agaccatcaa tcttgcacgc cgtacgcact ggggcggcga 14520
 cctgaaaacc atcctgcata ccaacatgc aaatgtgaac gagttcatgt ttaccaataa 14580
 gtttaaggcg cgggtgtatgg tttatgttgc ttggatccccc ttagtatgc 14640
 atacgacttgc gttggatccca cgtatgcgc gggcaactac tccgcacccgc tgacccataga 14700
 ccttatgaac aacgcgcgttcc tggagacta cttgaaatgtt ggcagacaga acgggggttct 14760
 ggaaaggcgcac atcggggtaa agtttgcaccc cgcgcacttc agactggggt ttgacccctg 14820
 cactggtctt gtcatgcctt gggatatac aaacgaagcc tccatccag acatcatttt 14880
 gctgcgcaggaa tgcgggggtgg acttcacccca cagccgcctt gacaacttgc tgggcattcc 14940

caagcggcaa cccttccagg agggcttag gatcacctac gatgatctgg	15000
cattccccca ctgttggatg tggacgccta ccaggcgagc ttgaaagatg	15060
gggcgggggt ggcgcaggcg gcagcaacag cagtggcage ggcgcggaaag	15120
cgcggcagcc gcccgaatgc agccgggtgg aacatgaac gatcatgcca	15180
caccttgc acacgggctg aggagaagcg cgctgaggcc gaagcagcgg	15240
cgcccccgct ggcgaaccccg aggtcgagaa gcctcagaag aaaccgggtga	15300
gacagaggac agcaagaaac gcagttacaa cctaataagc aatgacagca	15360
gtaccgcagc tggtaacctt catacaacta cggcgacccct cagaccggaa	15420
gaccctgtt tgcaactcctg acgtaaccctg cggctcgag caggtctact	15480
agacatgtg caagaccccg tgaccttccg ctccacgcgc cagatcagca	15540
ggtgtggcgcg gagctgttgc ccgtgcac tcacagatcc tacaacgcggc	15600
ctcccaactc atccggcagt ttacctctt gacccacgtg ttcaatcgct	15660
ccagattttg ggcgcggccgc cagccccccac catcaccacc gtcagtgaaa	15720
tctcacagat cacgggacgc taccgctcgcc caacagccatc ggaggaggtcc	15780
cattactgac ggcagacgcg gcacctgccc ctacgttac aaggccctgg	15840
gcccgcgtc ctatcgaccc gcaccccttg agcaagcatg tccatccta	15900
caataacaca ggctggggcc tgcgttccc aagcaagatg tttggcgggg	15960
ctccgaccaa caccctgtc ggcgtgcgcg gcactaccgcg gcccctggg	16020
acgcggccgc actgggcgcgca ccaccgtcgta tgacgcattc gacgcgggtgg	16080
gcccgcgtc ctatcgaccc gcacccacgc cgccaccagg gtccacagtg	16140
ggtgcgcgca gcccggcgct atgctaaaat gaagagacgg cggaggcgcg	16200
ccaccgcgcg cgcacccggca ctgcccggca acgcgcggcg gcccctgc	16260
acgtcgaccc ggccgacggg cggccatcgca ggccgcgtcg aggctggccg	16320
cactgtcccc cccaggtcca ggcgacgagc ggccgcggca gcagccgcgg	16380
tatgacttag ggtcgccagg gcaacgtgtt ttgggtgcgc gactcggtt	16440
cgtgcccgtg cgcacccggcc ccccgcgca cttagattgca agaaaaaaact	16500
gtactttgtt atgtatccag cggggggcggc ggcacacgaa gctatgttca	16560
caaagaagag atgctccagg tcatcgccgc ggagatctat ggccccccga	16620
gcaggattac aagccccgaa agctaaagcg gtcacaaaag aaaaagaaaag	16680
tgaacttgac gacgagggtgg aactgtgtca cgctaccgcg cccaggcgcac	16740
gaaaggtcga cgcgtaaaac gtgtttcgcc accggccacc accgtgtct	16800
tgagcgctcc accggccaccc acaagcgcgt gtatgttgcgt gttacggcg	16860
gctttagcgag gccaacgagc gcctcggggg gtttgcctac ggaaagcggc	16920
gctggcggtt cgcgtggacg agggcaaccc aacacccatgc ctaaagcccg	16980
gcagggtcg cccgcgcgtt caccgtccga agaaaagcgc ggccctaaagc	17040
tgacttggca cccaccgtgc agctgtatgtt acccaacgcgc cagcgtactgg	17100
ggaaaaaaatg accgtggAAC ctgggtctgg gcccgggttc cgcgtgcggc	17160
ggtggcgccg ggactggcg tgcagaccgt ggacgttcag atacccacta	17220
cagtattgcc accggccacag agggcatgg gacacaaacg tccccgggtt	17280
ggcggatgcc ggcgtgcagg cgggtcgctgc ggccgcgtcc aagacctcta	17340
aacggaccccg tggatgtttc gcgtttcgc ccccccggcgc cgcgcgggtt	17400
cggcgccggcc agcgcgtac tgccccgata tgccttacat cttccatttgc	17460
cggctatcggt ggctacaccc accggcccccag aagacgagca actacccgc	17520
cacttggaaacc cggccggccgcgtc gtgcgcacca ccggccggcg cgcgcgtcg	17580
cagggtggct cgcgaaggag gcaggaccctt ggtgtcgcca acagcgcgtt	17640
catcgtttAA aagccgttct ttgtggttct tgcagatatg gccttacactt	17700
tttccccgggtt cggggattcc gaggagaat gcaccgtagg agggcatgg	17760
cctgacgggc ggcacccgtc gtgcgcacca ccggccggcg cgcgcgtcg	17820
gcccggcggtt atccctggccc ttcttattcc actgtatcgcc gggcgatttgc	17880
cggaaatttgc tccgtggccct tgcaggcgca gagacactga taaaacaa	17940
gaaaaatcaa aataaaaaatgt ctggacttcc acgctcggtt ggtctgttac	18000
gaatggaaaga catcaactttt ggcgtctctgg ccccgacca cggctcgcc	18060
gaaactggca agatatcgcc accagcaata tgagcgggtgg cgccttcagc	18120
tgtggagcgg cattaaaaat ttcgggttcca ccgttaagaa ctatggcagc	18180
acagcagcac accggccacat cttggaggata agttgaaaga gcaaaaatttca	18240

tggtagatgg cctggcctct ggcattagcg gggtgtgga cctggccaac caggcagtgc	18300
aaaataagat taacagtaag cttgatecccc gcccctccgt agaggagccct ccaccggccg	18360
tggagacagt gtctccagag gggcgtggcg aaaagcgtcc ggcactaaag caaggcctgc	18420
ctctggtgac gcaaataagac gaggcctccct cgtacgagga ggcactaaag caaggcctgc	18480
ccaccaccccg tcccatcgcg cccatggcta cggagtgct gggccagcac acaccgtaa	18540
cgttggacct gcctccccc ggcgacaccc agcagaacc tgctgtgcca ggcccgaccg	18600
ccgttgttgt aaccgcgtt agccgcgcgt ccctgcgcg cggccgcgcgat ggtccgcgat	18660
cgttgcggcc cgtagccagt ggcaactggc aaagcacact gaacagcatc gtgggtctgg	18720
gggtgcaatc cctgaagcgc cgacgtatc tctgaatagc taacgtgtcg tatgtgttc	18780
atgtatgcgtt ccatcgccc gccagaggag ctgctgagcc gccgcgcgcg cgtttccaa	18840
gatggcttacc ctttcgatga tgccgcgtg gtcttacatg cacatctcg gccaggacgc	18900
ctcgaggatc ctgagccccc ggctggtgca ttggccgcg gccaccgaga cgtacttcag	18960
cctgaataac aagtttagaa accccacggg ggcccttacg cacgacgtga ccacagaccg	19020
gttcccggtt ttgacgctgc gtttcatccc tgggaccgt gaggatactg cgtactcgta	19080
caaggcgcgg ttccacccatg ctgtgggtga taaccgtgtg ctggacatgg cttccacgtt	19140
ctttgacatc cgcggcgtgc tggacagggg ccctacttt aagccctact ctggactgc	19200
ctacaacgccc ctggctccca agggtgcccc aaatcttc gatatggatg aagctgtac	19260
tgtcttggaa ataaacctag aagaagagga cgatgacaac gaagacgaag tagacgagca	19320
agctgagcag caaaaaactc acgtatggc gcaggcgcct tattctggta taaatattac	19380
aaaggagggtt attcaaatacg gtgtcgaagg tcaaacaactt aataatgcgc ataaaacatt	19440
tcaacctgaa cctcaaatacg gagaatctca gtggtaacaa actgaaatta atcatgcgc	19500
tgggagagtc ctaaaaaaga ctaccccaat gaaaccatgt tacgggttcat atgcaaaacc	19560
cacaaatgaa aatgggggc aaggcattct tggtaaagcaa caaaatggaa agctagaaag	19620
tcaagtggaa atgcaatttt tctcaactac tgaggcgcacc gcaggcaatg gtgataactt	19680
gactcctaaa gtggatttgt acagtgaaga tggatata gaaaccccgac acactcatat	19740
ttcttacatg cccactatta aggaaggtaa ctcacgagaa ctaatggcc aacaatctat	19800
gccccacagg cctaattaca ttgttttag ggacaatttt attggctaa tggattacaa	19860
cagcacgggtt aatatgggtt ttctggcgcc ccaagcatcg cagttgaatg ctgtttaga	19920
tttgcaagac agaaaacacag agctttcata ccagctttt cttgattcca ttgggtatag	19980
aaccaggatc ttttctatgt ggaatcaggc tggatcggc tatgatccag atgttagaat	20040
tattaaaaat catgaaactg aagatgaact tccaaattac tgctttccac tgggaggtgt	20100
gattaataca gagactctt ccaaggtaaa acctaaaaca ggtcaggaaa atggatggga	20160
aaaagatgtt acagaatttt cagataaaaaa tggaaataaga gtggaaata attttgcct	20220
ggaaatcaat ctaaatgcca acctgtggag aaatttctg tactccaaca tagcgctgt	20280
tttggccgac aagctaaagt acagtccttc caacgtaaaa attctgtata acccaaacac	20340
ctacgactac atgaacaatgc gagggtggc tccctgggtt gtggactgtc acattaacct	20400
tggagcgcgc tggcccttgc actatatggc caacgtcaac ccatttaacc accaccgcaa	20460
tgtctggctcg cgctaccgtt caatgttgc gggcaatgt cgctatgtgc cttccacat	20520
ccaggtgcct cagaaggttt ttgccattaa aaacccctt ctccgtccgg gctcatacac	20580
ctacgagtgg aacttcagga aggatgtttaa catgggtctg cagagctccc tagggaaatga	20640
cctaagggtt gacggagcca gcatggatgt tgatagcatt tgcccttacg ccacccctt	20700
ccccatggcc cacaacaccg cttccacgt tgaggccatg cttagaaacg acaccaacga	20760
ccagtcctt aacgactatc tccctggcgca acatgtctc taccctatac ccgccaacgc	20820
taccaacgtt cccatatcca tccctcccg caactggcg gcttccgcg gctggccctt	20880
cacgcgcctt aagactaagg aaaccccatc actgggtctg ggctacgacc cttattacac	20940
ctactctggc tctataccct acctagatgg aaccccttac ctaaccacaca ctttaagaa	21000
ggtgtggccatt acctttgact cttctgtca gtcggccgtt aatgaccgcg tgcttacccc	21060
caacgagttt gaaattaagc gtcagttga cggggagggt tacaacgtt cccagtgtaa	21120
catgaccaaa gactgggtcc tggatcaaatt gctagctaac tacaacattt gctaccagg	21180
tttctatatac ccagagagct acaaggaccg catgtactcc ttctttagaa acttccagcc	21240
catgagccgtt cagggtgggg atgataactaa atacaaggac taccacacagg tgggcaccc	21300
acaccaacac aacaactctg gatttgggg ctacccgttcc cccaccatgc gcgaaaggaca	21360
ggcctaccct gctaacttcc cctatccgct tataggcaag acccgagttt acagcattac	21420
ccagaaaaag tttctttggc atgcacccct ttggcgcaccc cattctcca gtaactttat	21480
gtccatgggc gcactcagac acctggccaa aacccctc tacgccaact ccggccacgc	21540

gctagacatg actttttaggg tggatcccat ggacgagccc acccttctt atgttttgtt 21600
 tgaagtctt gacgtggtcc gtgtgcaccc gccgcaccgc ggcgtcatcg aaaccgtgta 21660
 cctgcgcacg cccttctcg cccggcaacgc cacaacataa agaagcaagc aacatcaaca 21720
 acagctgcgg ccatgggctc cagttagcgag gaactgaaag ccattgtcaa agatcttggt 21780
 tgtggccat atttttggg cacctatgac aagcgcttcc caggctttgt ttctccacac 21840
 aagctcgctt ggcgcatacg caatacggcc ggtcgcgaga ctgggggctg acactggatg 21900
 gccttgcctt ggaacccgca ctcaaaaaca tgctacccctt ttgagccctt tggctttct 21960
 gaccagcgac tcaaggcagg ttaccagggt gactacgact cactctcgcc ccgtagcgcc 22020
 attgcttctt cccccgaccc ctgtataaagc ctggaaaatg ccacccaaag cgtacagggg 22080
 ccccaactcgcc cgccctgtgg actattctgc tgcatgttcc tccacgcctt tgccaactgg 22140
 ccccaaaactc ccatggatca caacccaccat atgaacctta ttacccgggtt acccaactcc 22200
 atgctcaaca gtccccagggt acagcccccctt ctgcgtcgca accaggaaca gctctacagc 22260
 ttccctggagc gccactcgcc ctacttccgc agccacaggcg cgcaggattt gaggccact 22320
 tcttttgtc acttgaaaaaa catgtaaaaaa taatgtacta gagacactt caataaaaggc 22380
 aaatgctttt atttgtacac ttcgggttga ttatccccc ccaccccttgc cgtctgcgc 22440
 gttttaaaat caaagggggtt ctgcccgcga tcgctatgcg ccactggcag ggacacgtt 22500
 cgatactggt gtttagtgc ccactttaaac tcaggcacaat ccatccgcgg cagctcggt 22560
 aagttttcac tccacaggct ggcgcaccatc accaacggctt ttagcaggc gggcgccat 22620
 atcttgaatg cgcagggtgg gcctccgcgc tgccgcgcg agttgcata cacagggtt 22680
 cagcacttggaa acactatcag cggccgggtgg tgccacgcctt ccagcacgtt cttgtcgag 22740
 atcagatccg cgtccagggtc ctccgcgtt ctcaggcacaat ccacccgcgg cagctcggt 22800
 tgccctccca aaaaggggcgc gtgcccaggc tttgagggtc acttcgcaccg tagtggcatc 22860
 aaaagggtgac cgtgcccggc ctggggcttgg ggatacagcg cctgcataaaa agccttgatc 22920
 tgcttaaaag ccacccgttgc ctttgcgcctt tcagagaaga acatgcgcga agacttgcgc 22980
 gaaaaactgtat tggccggaca ggccgcgtcg tgccacgcgc accttgcgtc ggtgttggag 23040
 atctgcacca catttggcc ccacccgttcc ttacacatct tggccttgc agactgtcc 23100
 ttccagcgcgc gtcgcccgtt ttcgcgtc acatccattt caatcacgtg ctcccttattt 23160
 atcataatgc ttccgtgttag acacttaagc tgccctcgatc ttcgcgtcgca gccgtgcagc 23220
 cacaacgcgc agcccggtgg ctgcgtatgc ttgttaggtca cctctgc当地 cgcactgcagg 23280
 tacgcctgc当地 ggaatcgccc catcatcgatc acaaagggtt tttgtgttgc gaaagggtc 23340
 tgcaacccgc ggtgtccctc gttcagccag gtcttcataa cggccgc当地 agcttccact 23400
 tggtcaggca gtagtttgc gttcgcctt agatcgatcc ccacccgttgc cttgtccatc 23460
 agcgcgcgc当地 cggccctccat gccccttc当地 cccgc当地 cgc当地 cggccactc 23520
 ttccatcaccg taatttactt ttcgcgttgc ctgggttcc cctcttc当地 ttgcgtccgc 23580
 ataccacgc当地 ccactgggtc gtcttcatttcc agccgc当地 cttgcgtc当地 acctcccttgc 23640
 ccacccgttgc当地 ttagcaccgg tgggttgc当地 aaacccacca ttgttaggc当地 cccatcttgc当地 23700
 ctcccttc当地 cgc当地 cccat gattacccctt ggtgtatggcg ggcgc当地 cggcc 23760
 gggcgttcc local ttttccctt gggcgc当地 agtcccaatccg cccgc当地 gaggt cgc当地 23820
 gggcgtgggtg tgccgc当地 cccgc当地 cggccgtc当地 tttgtatggcg ctccctc当地 ctcggactcg 23880
 atacgc当地 ccacccgttcc ttttggggcc gcccggggag gccgc当地 cggcc 23940
 gagc当地 acacgt cccatgggt tgggggacgt cccgc当地 cccgc当地 ctcgggggtg 24000
 gtttgcgtc gtc当地 cccactggcc atttccttcc cccataggca gaaaaagatc 24060
 atggagtc当地 tcgagaagaa gacagccctt accgc当地 cttgc当地 caccaccgc当地 24120
 tccaccgc当地 ccgc当地 acacgc当地 cccttaccacc ttccccc当地 cccgc当地 aggccccc当地 gcttgc当地 gaggt 24180
 gaggaagtgatc ttatcgc当地 gacccagggt tttgtatggcg aagacgc当地 ggaccgc当地 24240
 gtaccaacag aggataaaaaa gcaagaccag gacaacgc当地 aggccaaacgc当地 ggaacaacgc当地 24300
 gggcgggggg acgaaaggca tggc当地 gactac ctagatgtgg gagacgc当地 gcttgc当地 gaggt 24360
 catctgc当地 cccactggc当地 cccatctgc当地 gacgc当地 cccgc当地 aagacgc当地 cccatgc当地 cccgc当地 24420
 ctcgc当地 catag cggatgtc当地 cccatgc当地 gacgc当地 cccgc当地 tattctc当地 cccgc当地 cccatgc当地 24480
 cccaaacgc当地 aagaaaacgg cccatgc当地 cccgc当地 cccatgc当地 cccgc当地 cccatgc当地 24540
 tttgc当地 cccgtc当地 cccatgggtc当地 tccatcttcc cccatgggtc当地 cccatgggtc当地 24600
 ctatcctgc当地 gtc当地 cccatgggtc当地 cccatgggtc当地 cccatgggtc当地 cccatgggtc当地 24660
 gtc当地 catacctg atatcgc当地 cccatgggtc当地 cccatgggtc当地 cccatgggtc当地 24720
 gagc当地 gagaagc gccgc当地 gcaacgc当地 cccatgggtc当地 cccatgggtc当地 24780
 ggagtggttgg tggactc当地 gggc当地 gacccagggt cccatgggtc当地 cccatgggtc当地 24840

gaggtcaccc actttgccta cccggcactt aacctacccc ccaaggcat gggcacatc
atgagttagc tgatcgctcg ccgtgcgcag cccctggaga gggatgcaaa ttgcagaa
caaacagagg agggcctacc cgcaattggc gacgacgc tagcgcgcgt gcttcaa
cgcgacgctg ccgacttgg aagggcactt ggagcgacgc aaactaatga tggccgcagt gctcgattacc
gtggagctt ggtgcacatcg gcggttctt gctgacccgg agatgcacgc caagcttagag
gaaacatgc actacaccc tcgacaggc tacgtacgcc aggccctgcaaa gatctca
gtggagctt gcaacctgg ctccatcc ggaattttgc acgaaaaccg cttggccaa
aacgtctt attccacgc caagggcag ggcgcggcactacgtccg cgactgcgtt
tacttatttc tatgctacac ctggcagacg gccatggcg tttggcagca gtgcgttgag
gagtgcacc tcaaggagct gcagaaaactg ctaaagcaaa acttgaaggc cctatggac
gccttcaacg agcgtccgt ggccgcgcac ctggcggaca tcattttccc cgaacgcctg
cttaaaaccc tgcaacaggg tctgccagac ttccacagtc aaagcatgtt gcagaacttt
aggaacttta tcctagagcg cttaggaatc ttggccgcca cctgctgtgc acttcctagc
gactttgtgc ccattaagta ccgcgaatgc cttccggccg tttggggcca ctgctaccc
ctgcagctag ccaactaccc tcgctaccac tctgacataa tggaaagacgt gagcgggtgac
ggtctactgg agtgcactg tcgctgcaac ctatgcaccc cgcacccgctc cttgggttgc
aattcgcacg tgcttaacga aagtcaatt atcggtaccc ttgagctgca gggtcctcg
cctgacgaaa agtccgcggc tccgggggtt aaactcaactc cggggctgtg gacgtcggt
tacccgc aatttgtacc tgaggactac cacgcacccagc agattaggtt ctacgaagac
caatcccgc cgccaaatgc ggagcttacc gcctgcgtca ttacccaggg ccacatttt
ggccaattgc aagccatcaa caaagccgcg caagagttt tgctacgaaa gggacggggg
gtttacttgg accccccagtc cggcgaggag ctcaacccaa tccccccgccc gccgcagccc
tatcagcagc agccgcgggc cttgccttcc caggatggca cccaaaaaaga agctgcagct
ggccgcggca cccacggacg aggaggaata ctgggacagt caggcagagg aggttttgg
cgaggaggag gaggacatga tggaaagactg ggagagccta gacgagggaa cttccgaggt
cgaagagggt tcagacgaaa caccgtcacc ctgcgtcga ttcccctcg cggcgccccca
gaaatcgca accggttcca gcatggctac aaccccgct cctcaggcgc cggccggcact
ggccgttgc cgacccaaacc gtagatggga caccactggc accaggcccg gtaagtccaa
gcagccggcg ccgttagcccc aagagcaaca acagcccaaa ggctaccgc catggcgcgg
gcacaagaac gccatagttt ctgcgttgcg agactgtggg ggcaacatct cttcgcccc
ccgtttctt ctctaccatc acggcgtggc cttcccccgt aacatcctgc attactacc
tcatctctac agccataact gcaccggcgg cagcggcagc ggcagcaaca gcagccggca
cacagaagca aaggcgaccg gatagcaaga ctctgacaaa gccaagaaaa tccacagcgg
ccgcagcagc aggaggagga ggcgtgcgtc tggcccccgg cgaacccgtt tcgaccccg
agcttagaaa caggatttt cccactctgt atgctatatt tcaacagagc agggggcaag
aacaagagct gaaaataaaa aacaggtctc tgcgatccct caccggcagc tgccgttate
acaaaaagcga agatcagtt cggcgacgc tggaaagacgc ggaggctctc ttcagtaat
actgcgcgt gactttaag gactgtttt gcgcctttc tcaaattaa ggcgcggaaac
tacgtcatct ccagcggca caccggcgc cagcacctgt cgtcagcgc attatgagca
aggaaatcc cacggctac atgtggagtt accagccaca aatgggactt gcccgttgg
ctgccaaga ctactcaacc cgaataaaact acatgagcgc gggacccac atgatatccc
gggtcaacgg aatcccgccc caccgaaacc gaattttt ggaacaggcg gcttattacca
ccacaccccg taataaccc aatccccgtt gttggcccg tggccctgggt taccaggaaa
gtcccgctcc caccactgtg gtacttccca gagacgcccc gggcgaagtt cagatgacta
actcaggggc gcagcttgcg ggcggcttc gtcacagggt ggggtcgccc gggcagggt
taactcacct gacaatcaga gggcgaggta ttcaagcttca cgcacgactcg gtgagctt
cgcttgggtt ccgtccggac gggacattt agatcggcgg cggccggccgt ctttattca
cgccctgtca ggcaatccca actctgcaga cttcgcttcc tggccgcgc tctggaggca
ttggaaactt gcaattttt gaggagttt tgccatcggt ctactttaa cccttcctgg
gaccctcccg ccactatccg gatcaattt ttccatcc tgcacgcggta aaggactcgg
cgacggccta cgactgaatg ttaagtggag aggcagagca actgcgcctg aaacacctgg
tccactgtcg ccgcaccaag tgctttggc ggcactccgg tggattttgc tactttgaat
tgcccgagga tcataatcgag gggccggcgc acggcgctcg gcttaccgc cagggagagc
ttggccctgtccgatcc gatgttttcc aggcgcggccgt gcttagtttag gggacacagg
gaccctgtgt tctcaactgtg atttgcaact gtcctaaacct tggattacat caagatctt
24900
24960
25020
25080
25140
25200
25260
25320
25380
25440
25500
25560
25620
25680
25740
25800
25860
25920
25980
26040
26100
26160
26220
26280
26340
26400
26460
26520
26580
26640
26700
26760
26820
26880
26940
27000
27060
27120
27180
27240
27300
27360
27420
27480
27540
27600
27660
27720
27780
27840
27900
27960
28020
28080
28140

gttgcacatct	ctgtgctgag	tataataaaat	acagaaaatta	aatatactg	gggctccat	28200
cgccatcctg	taaacgccac	cgtcttcacc	cggccaaagca	aaccaaggcg	aaccttacct	28260
ggtacttttta	acatctctcc	ctctgtgatt	tacaacagt	tcaaccccaga	cggagtgagt	28320
ctacgagaga	acctctccga	gctcagctac	tccatcagaa	aaaacaccac	cctccttacc	28380
tgccggaaac	gtacgagtgc	gtcacccggcc	gctgcaccac	acctaccgccc	tgaccgtaaa	28440
ccagactttt	tccggacaga	cctcaataac	tctgttacc	agaacaggag	gtgagcttag	28500
aaaaccctta	gggttattagg	ccaaaggcg	agctactgtg	gggttatga	acaattcaag	28560
caactctacg	ggctattcta	attcagggtt	ctctagaatc	gggggtgggg	ttattctctg	28620
tcttgtgatt	ctcttttattc	ttataactaac	gcttctctgc	ctaaggctcg	ccgcctgtg	28680
tgtgcacatt	tgcattttt	gtcagctttt	taaacgtgg	ggtcgcccacc	caagatgatt	28740
aggtacataa	tccttaggtt	actcaccctt	gctgcagcccc	acggtaccac	ccaaaagggtg	28800
gatttttaagg	agccagcctg	taatgttaca	tgcgcagctg	aagctaata	gtgcaccact	28860
cttataaaat	gcaccacaga	acatgaaaag	ctgcttattc	gccacaaaaaa	caaaattggc	28920
aagtatgtcg	tttatgttat	ttggcagcca	ggtgacacta	cagagtataa	tgttacagtt	28980
ttccagggta	aaagtataaa	aactttatg	tatacttttc	cattttatga	aatgtgcac	29040
attaccatgt	acatgagcaa	acagtataag	tttgtggcccc	cacaaaattt	tgtggaaaac	29100
actggcactt	tctgcgtcac	tgctatgtct	attacagtgc	tcgctttgg	ctgtacccta	29160
ctctatattt	aataaaaaag	cagacgcagc	tttattgagg	aaaagaaaaat	gccttaattt	29220
actaagttac	aaagctaata	tcaccactaa	ctgcttact	cgctgcttgc	aaaacaaaatt	29280
caaaaaggta	gcattataat	tagaatagga	tttaaaccccc	ccggtcattt	cctgctcaat	29340
accattcccc	tgaacaattt	actctatgtg	ggatatgtc	cagcgtaca	accttgaagt	29400
caggcttct	ggatgtcagc	atctgacttt	ggccagcacc	tgtcccgccg	atttgttcca	29460
gtccaactac	agcgacccac	cctaaacagag	atgaccaaca	caaccaacgc	ggccgcccgt	29520
accggactta	catctaccac	aaatacaccc	caagttctg	cctttgtcaa	taactgggat	29580
aacttgggca	tgtgggggtt	ctccatagcg	cttatgttt	tatgccttat	tattatgtgg	29640
ctcatctgt	gectaaagcg	caaacgcgc	cgaccaccca	tctatagttc	catcattgtg	29700
ctacacccaa	acaatgtgg	aatccataga	ttggacggac	tggaaacacat	gttctttct	29760
cttacagttat	gattaaatga	gacatgattt	ctcgagttt	tatattactg	acccttgg	29820
cgctttttt	tgcgtgtcc	acattggctg	cggttctca	catcgaagta	gactgcattc	29880
cagccttac	agtctattt	cttacggat	ttgtcaccct	cacgctcatc	tgcaagctca	29940
tcactgtggt	catgcctttt	atccagtgc	ttgactgggt	ctgtgtgc	tttgcata	30000
tcagacacca	tccccagtag	agggacagga	ctatagctga	gettctttaga	attctttaat	30060
tatgaaattt	actgtgactt	ttctgtctat	tatttgcacc	ctatctgcgt	tttgttcccc	30120
gacctccaag	cctcaaagac	atatatcatg	cagatttact	cgatata	atattccaag	30180
ttgctacaat	aaaaaaagcg	atcttccga	agcctggta	tatgcata	tctctgttat	30240
ggtgttctgc	agtaccatct	tagccctag	tatataccc	taccttgc	ttggctggaa	30300
acgaatagat	gccatgaacc	acccaactt	ccccggcccc	gctatgttc	cactgcaaca	30360
agttgttgc	ggcggtttt	tcccagccaa	tcagcctc	ccacttctc	ccaccccccac	30420
tgaatcatgc	tactttaatc	taacaggagg	agatgactga	caccctagat	ctagaaatgg	30480
acggaattat	tacagagcag	cgcctgtctag	aaagacgcag	ggcagcggcc	gagcaacac	30540
gcatgaatca	agagctccaa	gacatggta	acttgaccca	gtgcaaaagg	gttatcttt	30600
gtctgtttaaa	gcaggccaaa	gtcacctac	acagtaata	caccggacac	cgccttagct	30660
acaagttgcc	aaccaagcgt	cagaatttgg	tggtcatgtt	gggagaaaaag	cccat	30720
taactcagca	ctcggtagaa	accgaaggct	gcatttactc	accttgc	ggacctgagg	30780
atctctgtcac	ccttattaa	accctgtcg	gtctcaaa	tcttattttt	tttactaat	30840
aaaaaaaaat	aataaaagcat	cacttactt	aatcgtt	gcaaaatttct	gtccagtt	30900
ttcagcagca	cctcccttgc	ctcttccca	ctctgtt	gcacgttct	cctggctgc	30960
aactttctcc	acaatctaaa	ttgaatgtca	gtttcttct	gttccctgtcc	atccgcaccc	31020
actatcttca	tgttgttgca	gatgaagcgc	gcaagacccgt	cttcaacccc	31080	
gtgtatccat	atgacacgg	accggctt	ccaaactgtc	cttttcttac	tcctccctt	31140
gtatccccca	atgggttca	agagagttcc	cctgggttac	tctttttgc	cctatccgaa	31200
cctctagttt	cctccaaatgg	catgttgc	ctcaaaaatgg	gcaacggcc	ctctctgg	31260
gaggccggca	accttac	ccaaaatgtt	accactgtga	gcccacctt	caaaaaaacc	31320
aagtcaaaaca	taaacctgga	aatatctgc	cccctcacag	ttacctcaga	agccctaact	31380
gtggctgc	ccgcacccct	aatggtgc	ggcaacacac	tcaccatgc	atcacaggcc	31440

ccgctaaacc tgcaacgactc caaacttagc attgccaccc aaggaccctt cacagtgtca 31500
gaaggaaagc tagccctgca aacatcaggc cccctcacca ccaccgatag cagtaccctt 31560
actatcactg ctcaccccc tctaactact gccactggta gcttgggcat tgacttgaaa 31620
gagcccattt atacacaaaa ttgaaaacta ggactaaagt acggggctcc tttgcatgta 31680
acagacgacc taaaacactt gaccgtagca actggccag gtgtgactat taataatact 31740
tccttgc当地 ctaaagttac ttggagcctt ggttttggatt cacaaggcaa tatgcaactt 31800
aatgttagcag gaggactaag gattgattt caaaacagac gccttataact tgatgttagt 31860
tatccgtttg atgctaaaa ccaactaaat ctaagactag gacaggcccc tcttttata 31920
aactcagccc acaacttgg tattaaactac aacaaaggcc ttacttgg tacagcttca 31980
aacaattcca aaaagcttga ggtaaaccta agcactgcca aggggttggat gtttgcgt 32040
acagccatag ccattaatgc aggagatggg cttaaattt gttcacctaa tgccacaaac 32100
acaaatcccc tcaaaacaaa aattggccat ggcctagaat ttgattgaaa caaggctatg 32160
gttccttaaac taggaactgg ccttagttt gacagcacag gtgcatttac agtaggaaac 32220
aaaaataatg ataagctaac ttgtggacc acaccagctc catctcttcaa ctgttagacta 32280
aatgcagaga aagatgctaa actcaactt gtcctaacaa aatgtggcag tcaaatactt 32340
gctacagttt cagtttggc tgtaaaggc agtttggctc caatatctgg aacaggctaa 32400
agtgc当地 tattataag atttgacgaa aatggagtgc tactaaacaa ttcttcctg 32460
gacccagaat attgaaacctt tagaaatggg gatcttactg aaggcacagc ctataacaaac 32520
gctgttgat ttatgcctaa cctatcagct tatccaaaat ctcacggtaa aactgccaaa 32580
agtaacattt tcagtc当地 ttacttacgg gtagacaaaaa ctaaacctgt aacactaacc 32640
attacactaa acggtagcaca gggaaacagga gacacaactc caagtgcata ctctatgtca 32700
ttttcatggg actggcttgg ccacaactac attaatgaaa tatttgcac atcctcttac 32760
acttttcat acattgcccc agaataaaaga atcggttggg ttatgttca acgtgtttat 32820
tttcaattt cagaaaaattt caagtcattt ttcatcagtt agtataccccc caccaccaca 32880
tagcttatac agatcaccgt accttaatca aactcacaga accctgtat tcaacactgccc 32940
acctccctcc caacacacag agtacacagt cctttctccc cggctggcc taaaaagcat 33000
catatcatgg gtaacagaca tattcttagg tgttatattc cacacggttt cctgtcgagc 33060
caaagctca tcagtgatata taataaactc cccggccagc tcacttaagt tcatgtcgct 33120
gtccagctgc tgagccacag gctgctgtcc aacttgcggg tgcttaacgg gggcgaagg 33180
agaagtccac gcctacatgg gggtagagtc ataatctgc atcaggataag ggcgggtgg 33240
ctgcagcagc ggc当地ataaa actgctggc cggccgc当地 gtcctgcagg aataacaacat 33300
ggcagtggtc ttctcagcga tgattcgcac cggccgcagc ataaggcgc ttgtccctcg 33360
ggcacagcagc cgccaccctga tctcaactt atcagcacag taactgc当地 acagcaccac 33420
aatattttc aaaatcccc accgtcaaggc gctgtatcca aagctcatgg cggggaccac 33480
agaacccacg tggccatcat accacaaggc cagtagatt aagtggcgc ccctcataaa 33540
cacgctggac ataaacatta cctcttttgg catgttggtaa ttccaccaccc cccggtagcca 33600
tataaaccctc tgattaaaca tggcccatc caccaccatc ctaaaaccaggc tggccaaaac 33660
ctgcccggc gctataact gcaggaaacc gggactggaa caatgacagt ggagagccca 33720
ggactctgaa ccatggatca tcatgctcgt catgatatac atgttggcac aacacaggca 33780
cacgtgcata cactcctca ggattacaag ctcccccgc gttagaaacca tatcccagg 33840
aacaacccat tcctgaatca gc当地aaatcc cacactgc当地 ggaagaccct gcacgttaact 33900
cacgttgc当地 attgtcaaaag ttgtacattt gggcagcagc ggatgatccct ccagtatgg 33960
agcggggtt tctgtctca aaggaggtag acgatccctt ctgtacggag tgcggccgaga 34020
caaccgagat cgttggc当地 gtatgtcat gccaaatggg acgccc当地 agtcatatt 34080
tcctgaagca aaaccaggc cggccgtgac aaacagatct gc当地ccgg tctcgccgct 34140
tagatcgctc tggtagtag ttgttagata tccactctt ctaaggcatcc aggccccc 34200
tggcttccgg ttctatgtaa actccttcat gccc当地tgc cctgataaca tccaccaccg 34260
cagaataagc cacacccaggc caacctacac attcgttctg cgagtc当地 acgggaggag 34320
cgggaaagagc tggaaagaacc atgtttttt ttttattcca aaagattatc caaaacctca 34380
aatgaagat ctattaagtg aacgc当地tcc cctccggg cgtggtaaaa ctctacagcc 34440
aaagaacaga taatggcatt tggtaagatgt tgccaaatgg cttccaaaag gcaaaacggcc 34500
ctcacgtcca agtggacgta aaggctaaac cttcagggt gaatctcttataa 34560
ccagcacctt caaccatgcc caaataattc tc当地ccgac accttctca tataatctca 34620
agcaaaatccc gaatattaag tccggccatt gtaaaaatct gtc当地agcc gccc当地acc 34680
ttcagctca agcagcgaat catgattgca aaaattcagg ttcttc当地ac acctgtataa 34740

gattcaaaaag	cggAACATTA	acaaaaatac	cgcgatcccg	taggtccctt	cgcaggGCCA	34800
gctgaacata	atcgTCcAGG	tctgcACGGA	ccAGCgCGGC	cacttccccG	ccAGGAACCT	34860
tgacaaaaga	acccACACTG	attatgACAC	gcataCTCGG	agctatGCTA	accAGCgtAG	34920
ccccgatgtA	agctttGTG	catgggCGGC	gataAAAT	gcaaggTGTc	gctcaaaaaA	34980
tcaggcaaaAG	cctcgcGCAA	aaaAGAAAGC	acatgtAGT	catgtCATG	cagataAAAG	35040
caggttaAGCT	ccggAACCCAC	cacAGAAAAA	gacACCATT	ttctCTCAAA	catgtCTCGG	35100
ggtttCTGCA	taaaACACAAA	ataaaATAAC	aaaaAAACAT	ttaaacATTA	gaAGCCTGTC	35160
ttacaACAGG	aaaaACAAACC	tttATAAGCA	taAGACGGAC	taCGGCCATG	ccggCGTGAC	35220
cgtaaaaaaaa	ctggTCACCG	tgatTTAAA	gcaccACCGA	cagtCTCTCG	gtcatgtCCG	35280
gagtCATAAT	gtaAGACTCG	gtAAACACAT	caggTTGATT	catcggtCAG	tgctaaaaAG	35340
cgaccGAAAT	agccccGGGG	aatacataCC	cgcaggCGTA	gagacaACAT	tacAGCCCCC	35400
ataggAGGTA	taacAAAATT	aataggAGAG	aaaaACACAT	aaACACCTGA	aaaACCCCTC	35460
tgccttaggCA	aaatAGCACC	ctcccGCTCC	agaACAAACAT	acAGCGCTTC	acAGCGGCAg	35520
cctaACAGTC	agcTTACCA	gtaaaaAAAGA	aaACCTTAA	aaaaAAACACC	actcgACACG	35580
gcaccAGCTC	aatcAGTCAC	agtgtAAAAA	agggCCAAGT	gcAGAGCAG	tatataTAGG	35640
actaaaaAAAT	gacgtAACGG	ttaaAGTCC	caaaaaAAACAC	ccagaaaaAC	gcacgcGAAC	35700
ctacGCCCAg	aaACGAAAGC	caaaaaACCC	acaACTTCCT	caaATCGTCA	cttccGTTT	35760
cccacGTTAC	gtaACTTCCC	atTTAAGAA	aactACAATT	cccaACACAT	acaAGTTACT	35820
ccgcCCTAAA	acctACGTCA	cccgCCCCGT	tcccACGCC	cgcGCCACGT	cacaAACTCC	35880
acccCCTCAT	tatCATATTG	gCTTCATCC	aaaATAAGGT	atattATTGA	tgatG	35935

<210> 10
<211> 5965
<212> DNA
<213> Artificial Sequence

<220>
<223> NSsuboptmut

<400> 10	60					
gccaccatgg	cccccatcac	cgcctacAGC	cagcAGACCA	ggggCCTGCT	gggCTGCATC	60
atcaccAGCC	tgaccGGACG	cgacaAGAAC	caggTTGGAGG	gagAGGTGCA	ggTGGTgAGC	120
accgCTACCC	agAGCTTCT	ggccACCTGc	gtGAACGGCG	tgtGCTGGAC	cgtGTACCAc	180
ggagCCGGAA	gcaAGACCT	ggCCGGACCC	aaggGCCtTA	tcACCCAGAT	gtACACCAAT	240
gtggatCAGG	atCTGGTGGG	ctggcAGGCC	cctCCCGAG	ccaggAGCCT	gacACCCtGT	300
acCTGTGGAA	gcAGCGACCT	gtACCTGGT	acACGCCAG	ccGATGTGAT	ccccGTGAGG	360
cgCAGGGCG	attCTCGGG	aAGCCTGCTG	agCCCTAGGC	ccGTGAGCTA	cctGAAGGGC	420
agcAGCggAG	gACCCtGTCT	gtgtCETTCT	ggCCATGCCG	tggGCATTtT	tcgcGCTGCC	480
gtgtGTACCA	ggggCgtGGC	caaAGCCTG	gatTTGTGc	ccGTGGAAG	catGGAGACC	540
accatGCGCA	gcccTGTGTT	caccGACAAAC	agCTCTCCCC	ctGCCGTGCC	ccaATCATTc	600
caggTGGTC	acCTGcACGC	ccCTACCGGA	tctGGCAAGA	gcACCAAGGT	gccccGCTGCC	660
tacGCCGTC	aggGCTACAA	gtGTCTGGT	ctGAACCCCA	gcgtGGCCGc	taccCTGGC	720
ttcggcGCTT	acATGAGCAA	ggcccATGGC	atcgACCCCA	acATCCGcAC	aggcGTCGc	780
accatCACCA	ccggAGCTCC	ctgtGACCTAC	agCACCtACG	gcaAGTTCT	ggccGATGGA	840
ggCTGcAGCG	gaggAGCCTA	cgACATCAtC	atCTGcACG	atGcCCACAG	caccGACAGC	900
accACCACTC	tggcATTGG	caccGTGCTG	gatCAGGCCG	aaACAGCTGG	agCCAGGCTG	960
gtggTGTCTG	ccACAGCTAC	ccCTCCTGGC	agcGTGACCG	tGCCCCATCC	caATATCGAG	1020
gaggGTGGCC	tgAGCAACAC	aggcGAGATC	ccCTTCTACG	gcaAGGCCAT	ccccATCGAG	1080
gCCATCCGCG	gaggCAGGCA	cctGATCTC	tGCCACAGCA	agaAGAAGTG	cgACGAGCTG	1140
gCTGCCAAGC	tgAGCGGACT	ggGCATCAAC	gCCGTGGCT	actACAGGGG	cctGGACGTG	1200
tcAGTGTATCC	ccACCATCGG	cgATGTGGT	gtGGTGGCA	ccGACGCCCT	gatGACAGGC	1260
tacACCGAG	actTCGACAG	ctgtGATCGAC	tgCAACACCT	gctGTGACCC	gaccGTGGAC	1320
ttcAGCCTGG	acCCACCTT	caccATCGAA	accACCCACG	tGCCCTAGGA	tgCTGTGAGC	1380
aggAGCCAGA	ggcGCGGACG	caccGGAAAGG	ggcAGGCCG	gaATTtatCG	ctttGTGACC	1440
cctGGCGAAA	ggccCTCTGG	catgtTCGAC	agcAGCGTGC	tgtGCGAGTG	ctacGACGCT	1500

ggctgcgtt	ggta	cgcgt	gacacccgt	gaaaccagcg	tgcgcctgcg	cgcttatctg	1560
aataccccgt	gcctgcccgt	gtgtcaggac	cacctggagt	tctgggagag	cgtttcac		1620
ggactgaccc	acatcgacgc	ccatttcctg	agccagacca	agcaggctgg	cgacaacttc		1680
ccctatctgg	tggcttatca	ggccaccgtg	tgtcttaggg	cccaagctcc	acctccttca		1740
tgggaccaga	tgttgaagtg	cctgatccgc	ctgaagccca	cctgtcacgg	ccctaccct		1800
ctgctgtacc	gcctgggagc	cgtcagaac	gaggtgaccc	tgaccccaccc	catcaccaag		1860
tacatcatgg	cctgcgttgc	cgtgtatctg	gaagtggta	ccagcacctg	ggtgcgttg		1920
ggaggcgtgc	tggccgtct	ggctgcctac	tgcctgacca	ccggaagcgt	ggtgatcgtg		1980
ggacgcatac	tectgagcgg	aggcccgt	atcgtcccc	atcgcgagtt	cctgtaccag		2040
gagttcgaac	agatggagga	gtgtgccagc	cacctgcctt	acatcgagca	gggcatgcag		2100
ctggccgaac	agttcaagca	gaaggccctg	ggcctgtgc	agacagccac	caaacaggcc		2160
gaagctgcgc	ctcccgttgt	ggaaaagcaag	tggagggccc	tgagacctt	ctgggctaag		2220
cacatgttga	acttcatctc	tggcatccag	tacctggccg	gactgagcac	cctgccttgc		2280
aaccccgcta	tcgcgcgtt	gatggccttc	accgcgtac	tcaacctctcc	cctgaccacc		2340
cagagcaccc	tgctgttcaa	cattctgggc	ggatgggtgg	ccgcgtcagct	ggcccttcct		2400
tcagctgttt	ctgcctttgt	gggcgtctggc	attgcggag	ccgcgtgtgg	cagcattggc		2460
ctgggcaaaag	tgctgttgga	tattctggct	ggctatggcg	ctggcgtggc	cggagccctg		2520
gtggccttca	aggtgtatgag	cgagagatg	cccagcaccc	aggacctgg	gaacctgtg		2580
cctgcccattc	tgagccctgg	agccctgggt	gtgggggtgg	tgtgtgtgc	cattctgagg		2640
cgccatgtgg	gaccggaga	gggcgtgtg	cagtggatga	accgcctgtat	cgcccttcgccc		2700
tctcgcgaaa	accacgttag	ccctacccac	tacgtgcctg	agagcgacgc	cgctgcccagg		2760
gtgacccaga	tcctgagcag	cctgaccatc	acccagctgc	tgaagcgcct	gcaccagtgg		2820
atcaacgagg	actgcagcac	accctgcagc	ggaagctggc	tgagggacgt	gtgggactgg		2880
atctgcaccc	tgctgaccga	tttcaagacc	tggctgcaga	gcaagctgtct	gcccccaactg		2940
cctggcgtgc	ccttcttctc	atgcagcgc	ggatacaagg	gctgtgtggag	gggcgtatggc		3000
atcatgcaga	ccacccgttcc	ctgcggagcc	cagatcacag	gccacgtgaa	gaacggcagc		3060
atgcgcac	tggcccttaa	gacctgcage	aacacccgttgc	acggcacctt	ccccatcaac		3120
gcctacacca	ccggaccctg	cacacccagc	cctgctccca	actacagcag	ggccctgtgg		3180
agggtggctg	ccgaggagta	cgtggaggtg	accagggtgg	gagacttcca	ctacgtgacc		3240
ggaatgacca	ccgacaacgt	gaagtgtccc	tgtcaggtgc	ccgcctccga	attttttacc		3300
gaagtggatg	gcgtgcgcct	gatcgctat	gcccctgcct	gtaggccct	gctgcgcgaa		3360
gaagtgaccc	tccaggtggg	cctgaaccag	tacctgggtgg	gcagccagct	gcccctgcgag		3420
cctgagcccg	atgtggccgt	gctgaccagg	atgctgaccg	accccagcca	catcacagcc		3480
gaaaccgcata	aaaggccct	ggccagggggc	tccctccaa	gctggccctc	aagcagcgct		3540
agccagctgt	ctgctccca	cctgaaggcc	acctgcacca	ccaccacgt	gagccccgac		3600
gccgaccctg	tcgaggccaa	cctgctgtgg	cgccaggaga	tggcggcaa	catcacccgc		3660
gtggagagcg	agaacaagggt	ggtgtgtctg	gacagcttcg	acccctgcg	cgccgaggag		3720
gacgagcgcg	aggtgagcgt	ccccggcag	atcctgcgc	agagcaagaa	gttccccgt		3780
gccatgccc	tctggcttag	acctgattac	aaccctcccc	tgtggagag	ctggaaaggac		3840
cctgattacg	tgcctccagt	ggtgcattgc	tgcctctgc	ctccctttaa	agccccctcc		3900
attccaccc	ctaggcgc	aaaggaccgt	tgtctgacag	aaagcagcgt	gagctctgt		3960
ctggccgaac	tggccaccaa	gacctttggc	agcagcgaga	gctctgcgt	ggacagcgga		4020
acagccaccc	ctctgcctg	ccaggccagc	gacgacggcg	ataagggcag	cgatgtggag		4080
agctatagca	gatgcctcc	cctgaaaggc	gaaacctggcg	atcccgatct	gagcgatggc		4140
agctggagca	ccgtgagcga	agaggccagc	gaggacgtgg	tgtgttgcag	catgagctac		4200
acctggacag	gctgcgtat	cacaccc	gctgcgcagg	agagcaagct	gccccatcaac		4260
gccctgagca	acagcctgt	gaggcaccac	aacatgggt	acgcaccac	cagcaggct		4320
gccggactg	ggcagaagaa	ggtgaccc	gaccgcctgc	aggtgctgg	cgaccactac		4380
cgcgatgtgc	tgaaggagat	gaaggccaa	gcccgcaccc	tgaaggccaa	gctgctgagc		4440
gtggaggagg	cctgcaagct	gacccccc	cacagcgc	agagcaagtt	cggtacggc		4500
gccaaggacg	tgcgcaac	gagcagcaag	gccgtgaacc	acatccacag	cgtgtggaa		4560
gacctgctgg	aggacaccgt	gacccccc	gacaccacca	tcatggccaa	gaacgagggt		4620
ttctgcgtgc	agccccgagaa	gggcggccgc	aagccgc	gcctgatcgt	gttccccgt		4680
ctgggcgtgc	gcgtgtgc	gaagatggcc	ctgtacgac	tggttagc	cctgcctcag		4740
gtgggtatgg	gctcaagcta	cggctccag	tacagccctg	gccagcgct	ggagttccctg		4800

gtgaacacct	ggaagagcaa	gaagaacccc	atggc ttca	gctacgacac	acgctgc ttc	4860
gacagcaccc	tgaccgagaa	cgacatccgc	gtggaggaga	gcatctacca	gtgctgc gac	4920
ctggccccctg	aggccaggca	ggccatcaag	agcctgaccg	agcgc ctgt	catcg gagc	4980
cctctgacca	acaga caagg	acagaactgc	ggatacaggc	gctgtagg	gc ctggc gt	5040
ctgaccacca	gctgtggcaa	caccctgacc	tgctac	aggccagc gc	tgc ctgtc	5100
gctgccaagc	tgcaggactg	caccatgctg	gtgaacgc	ctggc ctgtt	ggtgattt	5160
gaaagcgc	gcacccagg	agatgctgc	agcctgc	tgttac	ggccatgacc	5220
aggta	ctg cccccc	agacccccc	cagcc	gat	acccctgga	5280
agctgctcaa	gcaacgtg	cgtggctc	gacgc	gaaagcgc	gtactac	5340
acacgcgatc	ccacca	cccc	tctggc	gctgc	ccat	5400
gtgaacag	ctggggcaa	catcat	catgc	ccctgt	tcgc	5460
ctgatgaccc	actt	ttc	catc	ctgc	gatc	5520
tgccagat	acggc	gtt	ctac	acgc	gag	5580
cgcc	ctgc	act	acgc	gag	ccat	5640
gtggcc	act	ggc	act	ggc	atcg	5700
aggagcgt	ggg	ctagg	ctg	gac	gggg	5760
ctgttca	act	ggg	act	ggg	ccat	5820
ctggac	act	ggg	act	ggg	atcg	5880
tctc	cg	cc	ctgc	ctgc	ccac	5940
atctac	cc	cc	atgc	atgc	atcg	5965
at	cc	cc	atgc	atgc	atcg	

<210> 11
<211> 5965
<212> DNA
<213> Artificial Sequence

<220>
<223> Chimeric NSsuboptmut

<400> 11	60												
gccaccatgg	ccccatcac	cgcc	tacaggc	cagc	gaccc	gccc	cctg	ctgc	ggg	ctgc	atc	60	
atcaccagcc	tgaccggcc	cgaca	aga	aac	cagg	tgagg	g	cgagg	gtc	gtt	ggg	atc	120
accgc	cccc	agat	ttc	cct	ggcc	ac	gt	gac	gg	ctg	gg	ac	180
ggc	ccgg	ca	gaa	ccct	ggcc	gg	gg	cc	gg	gg	gg	cc	240
gtggacc	agg	ac	ctt	gtt	gg	ctt	gg	cc	gg	cc	gg	cc	300
ac	ctt	gg	gg	cc	tt	gg	gg	cc	gg	gg	gg	gg	360
cc	cc	gg	gg	gg	cc	gg	gg	gg	gg	gg	gg	gg	420
agc	agc	gg	gg	gg	cc	gg	gg	gg	gg	gg	gg	gg	480
gtgt	gac	cc	cc	gg	cc	gg	cc	gg	cc	gg	cc	gg	540
accat	gc	cc	cc	gtt	cc	gg	ac	ac	cc	cc	gg	cc	600
cagg	tgg	cc	ac	tgc	cc	cc	ac	cc	cc	cc	cc	cc	660
ta	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	720
ttc	ggc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	780
acc	at	at	at	at	at	at	at	at	at	at	at	at	840
gg	ct	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	900
acc	acc	at	at	at	at	at	at	at	at	at	at	at	960
gt	gg	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1020
gag	gt	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1080
g	cc	at	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1140
g	cc	at	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1200
gg	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1260
cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1320
cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1380
cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1440
cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	1500

ggctgcgcct	ggtacgagct	gaccccccgc	gagaccagcg	tgcgcctgcg	cgcctacctg	1560
aacacccccc	gcctgcccgt	gtgccaggac	cacctggagt	tctgggagag	cgtttcacc	1620
ggcctgaccc	acatcgacgc	ccacttctcg	agccagacca	agcaggccgg	cgacaacttc	1680
ccctacctgg	tggcttacca	ggccacccgtg	tgcgcctgcg	cccaggcccc	ccccccccc	1740
tgggaccaga	tgttgaagtg	cctgatccgc	ctgaagccca	ccctgcacgg	ccccaccccc	1800
ctgtgttacc	gcctggcgc	cgtgcagaac	gaggtgaccc	tgacccaccc	catcaccaag	1860
tacatcatgg	cctgcatgag	cgccgcacccgt	gaggtggta	ccagcacctg	ggtgctggtg	1920
ggccgcgtgc	tggccgcct	ggccgcctac	tgcctgacca	ccggcagcgt	ggtgatcgtg	1980
ggccgcatcc	tcctgagccg	ccgcggccgc	atcgtccccg	accgcgagtt	cctgtaccag	2040
gagttcgacg	agatggagga	gtgcgcgc	cacccgcct	acatcgagca	gggcacatgcag	2100
ctggccgcgc	agttaagca	gaaggccctg	ggccctgtgc	agaccgcac	caagcaggcc	2160
gaggccgcgc	ccccctgtgt	ggagagcaag	tggcgcgc	tggagacctt	ctgggccaag	2220
cacatgttgg	acttcatcag	cggcacatccag	tacctggccg	gcctgagcac	cctgcccggc	2280
aaccccgcca	tcgcacgcct	gatggccttc	accgcacca	tcaccagccc	cctgaccacc	2340
cagagcaccc	tgctgttcaa	catcctgggc	ggctgggtgg	ccgcggccgt	ggcccccccc	2400
agcgcgcaca	gcccgcgtgt	gggcgcgcgc	atgcgcggc	ccgcgtgggg	cagcatcgcc	2460
ctgggcaagg	tgctgttgg	catcctggcc	ggctacggcg	ccggcgtggc	cgccgcgcctg	2520
gtggccttca	aggtgatgag	cgccgcacccgt	cccacccgt	aggacccgtgt	gaacctgtgt	2580
cccgccatcc	tgagcccccgg	cgccctgggt	gtgggcgtgg	tgtgcgcgc	catcctgcgc	2640
cgccacgtgg	gccccggcga	gggcgcgcgt	cagtggatga	accgcctgtat	cgcccttcgc	2700
agccgcggca	accacgttag	ccccacccac	tacgtccccg	agagcgacgc	cgccgcgcgc	2760
gtgacccaga	tcctgagccg	cctgaccatc	accacgtgc	tgaagcgcct	gcaccagtgg	2820
atcaacgagg	actgcacgcac	ccctgcacgc	ggcagctggc	tgcgcgcacgt	gtgggactgg	2880
atctgcaccc	tgctgacccg	ttcaagacc	tggctgcaga	gcaagctgt	gccccagctg	2940
cccgccgtgc	ccttcttcag	ctgcacgcgc	ggctacaagg	gcgtgtggcg	cgccgcacggc	3000
atcatgcaga	ccacccgc	ctgcggcgc	cagatcaccg	gccacgtgaa	gaacggcagc	3060
atgcgcacgt	tggccccc	gacccgcacgc	aacacccgtgc	acggcacctt	ccccatcaac	3120
gcctacacca	ccggccctg	caccccgac	ccgcggccca	actacagccg	cgccctgtgg	3180
cgcggtggcc	ccgaggagta	ctggagggtg	acccgcgtgg	gcgacttcca	ctacgtgacc	3240
ggcatgacca	ccgacaaacgt	gaagtgc	tgccaggtgc	ccgcggccga	gttccatcacc	3300
gaggtggacg	gcgtgcgcct	gcacccgtac	gccccccgc	gcgcggcc	3360	
gaggtgaccc	tccaggtgg	cctgaacc	tacctgggtgg	gcagccagct	gcctgcgc	3420
cccgagcccg	acgtggccgt	gtgacccgc	atgctgaccc	acccacccca	cateacccgc	3480
gagaccgcca	agcgcgcgc	ggccgcgcgc	agccccccca	gcctggccag	cagcagcgc	3540
agccagctga	gcccgcgc	cctgaaggcc	acctgcacca	ccacccacgt	gagccccgac	3600
gccgacctga	tcgaggccaa	cctgcgtgtgg	cgccaggaga	tggcggccaa	catcaccgc	3660
gtggagagcg	agaacaagg	gttgggtgt	gacagcttcg	acccctgcg	cgccgaggag	3720
gacgagcgc	aggtgacccgt	ggccgcgc	atcctgcgc	agagcaagaa	gttcccccgt	3780
gccatgcaca	tctggcttag	acctgattac	aaccctcccc	tgcctggagag	ctggaaaggac	3840
cctgattacg	tgcttccagt	ggtgcacggc	tgcctctgc	ctccctattaa	agccccctcc	3900
attccaccc	ctaggcgca	aaggaccgtg	tgcctgacag	aaagcagcgt	gagctctgt	3960
ctggccgaac	tggccaccaa	gaccccttgc	agcagcgc	gctctgcgt	ggacagcgg	4020
acagccaccc	ctctgcctga	ccaggccacgc	gacgacggcg	ataagggcag	cgatgtggag	4080
agctatagca	gcatgcctcc	cctggaaaggc	gacccctggcg	atcccgatct	gagcgatggc	4140
agctggagca	ccgtgacgc	agaggccacgc	gaggacgtgg	tgtgttgcag	catgagctac	4200
acctggacag	gcgcgtctgt	cacccctgc	gctgcgc	aggaccaagct	ccccatcaac	4260
gccctgagca	acagccgt	gaggcaccac	aacatgggt	acgcacccac	cagcaggct	4320
gccggactga	ggcagaagaa	ggtgaccc	gaccgcctgc	agggtgtgg	cgaccactac	4380
cgcgatgtgc	tgaaggagat	gaaggccaa	gcccaccc	tgaaggccaa	gctgtgc	4440
gtggaggagg	cctgcaagct	gacccccc	cacagccca	agagcaagtt	cggttacggc	4500
gccaaggacg	tgcgcaccc	gagcagcaag	gccgtgaacc	acatccacag	cggttggag	4560
gacccgttgg	aggacaccgt	gacccccc	gacaccacca	tcatggccaa	gaacgaggt	4620
ttctgcgtgc	agcccagagaa	gggcggccgc	aagccccc	gcctgatcgt	gttcccccac	4680
ctgggcgtgc	gcgtgtgc	gaagatggcc	ctgtacacg	tggtggac	cctgccccag	4740
gtgggtatgg	gcagcagcta	cggttccag	tacagccccg	gccagcgcgt	ggagttcc	4800

gtgaacacct	ggaagagcaa	gaagaacccc	atggcgttca	gctacgacac	ccgctgcgttc	4860
gacagcaccc	tgaccgagaa	cgacatccgc	gtggaggaga	gcatctacca	gtgctgcgac	4920
ctggcccccc	aggcccccca	gccatcaag	agcctgaccg	agcgccctgt	catcgccggc	4980
ccccctgacca	acagcaaggg	ccagaactgc	ggctaccgccc	gctgcccgcgc	cagcggcgtg	5040
ctgaccacca	gctgcggcaa	caccctgacc	tgctactcta	aggccagcgc	ccgctgcgc	5100
gccgccaagc	tgcaggactg	caccatgctg	gtgaacgcgg	ccggccttgtt	ggtgatctgc	5160
gagagcgcgg	gcacccagga	ggacgcgcgc	agcctgcgcg	tgttacccga	ggccatgacc	5220
cgctacagcg	cccccccccgg	cgaccccccc	cagccccagg	acgaccttgg	gctgatcacc	5280
agctgcagca	gcaacgtgag	cgtggcccac	gacgcccagcg	gcaagcgcgt	gtactacctg	5340
acccgcgacc	ccacccacccc	cctggcccg	ggccgcctgg	agacgcgcgc	ccacaccccc	5400
gtgaacagct	ggctggccaa	catcatcatg	tacgcgcgc	ccctgtgggc	ccgcgcgcgc	5460
ctgatgaccc	acttcttcag	catcctgcgt	gcccaggagc	agctggagaa	ggccctggac	5520
tgccagatct	acgggcgcgt	ctacagcattc	gagccccctgg	acctgcgcgc	gatcatcgag	5580
ccgctgcacg	gctgagcgc	tttcagcctg	cacagctaca	gccccggcga	gatcaacccgc	5640
gtggccagct	gcctgcgcaa	gctggcgtg	ccccccctgc	gctgtggcg	ccacccgcgc	5700
cgcagcgtgc	gcccgcct	gctgagccag	ggcggccgcgc	ccgcccacctg	cgccaaatgc	5760
ctgttcaact	ggccgtgaa	gaccaagctg	aagctgaccc	ccatccccgc	cgccagccag	5820
ctggacctga	gccccgtgtt	cgtggccgc	tacagcggcg	gacacatcta	ccacagcctg	5880
agccgcgcgc	gccccgcgt	gttcatgcgt	tgccctgcgt	tgctgagcgt	ggcgtggc	5940
atctacctgc	tgcccaacccg	ctaaa				5965

<210> 12

<211> 10

<212> RNA

<213> Artificial Sequence

<220>

<223> Ribosome binding site

<400> 12

gccaccaugg

10

<210> 13

<211> 49

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic polyadenylation signal

<400> 13

aaauaaaagau cuuuauuuuc auuagaucug uguguugguu uuuugugug

49

<210> 14

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Additional nucleotides present in pVIJns-NS

<400> 14

tcttagagcgt ttaaaccctt aattaagg

28

<210> 15

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Additional nucleotides present in pV1Jns-NSOPTmut

<400> 15
tttaaatgtt taaac 15

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide primer

<400> 16
tcgaatcgat acgcgAACCT acgc 24

<210> 17
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide primer

<400> 17
tcgacgtgtc gacttcgaag cgcacaccaa aaacgtc 37

THIS PAGE BLANK (USPTO)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 April 2003 (17.04.2003)

PCT

(10) International Publication Number
WO 03/031588 A3

(51) International Patent Classification⁷: **C12N 15/40,**
15/51, 15/85, 15/86, 15/861, A61K 48/00

[IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT).
COLLOCA, Stefano [IT/IT]; Via Pontina KM. 30.600,
I-00040 Pomezia (IT).

(21) International Application Number: **PCT/US02/32512**

(74) Common Representative: **MERCK & CO., INC.**; 126
East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(22) International Filing Date: 10 October 2002 (10.10.2002)

(25) Filing Language: English

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC,
VN, YU, ZA, ZM, ZW.

(26) Publication Language: English

(30) Priority Data:
60/328,655 11 October 2001 (11.10.2001) US
60/363,774 13 March 2002 (13.03.2002) US

(71) Applicants (*for all designated States except US*): **MERCK & CO., INC.** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI, S.P.A.** [IT/IT]; VIA PONTINA KM. 30.600, I-00040 POMEZIA (IT).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **EMINI, Emilio**, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **KASLOW, David, C.** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **BETT, Andrew, J.** [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **SHIVER, John, W.** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **NICOSIA, Alfredo** [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). **LAHm, Armin** [DE/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). **LUZZAGO, Alessandra** [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). **CORTESE, Riccardo**

(76) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BI, BJ, CI, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
30 October 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 03/031588 A3

(54) Title: HEPATITIS C VIRUS VACCINE

(57) Abstract: The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell-mediated immune (CMI) response against HCV.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/32512

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/40, 15/51, 15/85, 15/86, 15/861; A61K 48/00
 US CL : 514/44; 424/93.2; 435/320.1, 455, 456

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 U.S. : 514/44; 424/93.2; 435/320.1, 455, 456

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,127,116 A (RICE et al.) 03 October 2000 (03.10.2000), column 45, lines 18-57.	1, 2
A	WO 01/30812 A2 (CHIRON CORPORATION) 03 May 2001 (03.05.2001).	1-54
A	WO 97/47358 A1 (MERCK & CO., INC.) 18 December 1997 (18.12.1997).	1-54

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

09 July 2003 (09.07.2003)

Date of mailing of the international search report

02 SEP 2003

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
 Commissioner for Patents
 P. O. Box 1450
 Alexandria, Virginia 22313-1450
 Facsimile No. (703)305-3230

Authorized officer

Scott D. Priebe

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/32512

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-54

Remark on Protest The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/32512

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-54, drawn to a nucleic acid encoding a HCV polyprotein.

Group II, claim(s) 55-59, drawn to a chimeric adenovirus vector comprising sequence derived from human adenovirus serotypes 5 and 6.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The technical feature of invention I is a nucleic acid encoding a polyprotein derived from an HCV polyprotein, whereas the technical feature of invention II is a chimeric adenoviral vector comprising a heterologous sequence. These two features are not related. Invention I does not require vector of invention II, nor does is the vector of invention II required to contain the polynucleotides of invention I.

Continuation of B. FIELDS SEARCHED Item 3:

MEDLINE, EMBASE, CAPLUS, BIOSIS, SCISEARCH, USPT, PGPB, DERWENT, GENBANK, GENESEQ
search terms: HCV, hepatitis C virus, vaccine, NSSB, NSSB near inactiv? or non-functional, SEQ ID NO: 1, SEQ ID NO: 2